CITY UNIVERSITY OF HONG KONG

Course code & title : CS3402 Database Systems

Session : Semester A 2019/20

This paper has 13 pages (including this cover page).

---

1. This paper consists of Five questions.

2. Write down your answer in the space provided.

---

*This is an **open-notes** examination.*

Candidates are allowed to use the following materials/aids:

Book,printed lecture notes, personal notes, and other course handout materials.

Seat Number	
Student ID	

Q1 (10%)	Q2 (24%)	Q3 (30%)	Q4 (16%)	Q5 (20%)	Total (100%)

Question 1 (10 marks): Basic RDBMS Concepts

Determine whether each of the following statements is true or false and justify your answers. No mark will be given if your justification is wrong.

(a) (2 marks): To perform a binary search on a relation, we first need to build a B+-tree.

(b) (2 marks): To process any SQL query with a WHERE clause, selection has to be performed before projection.

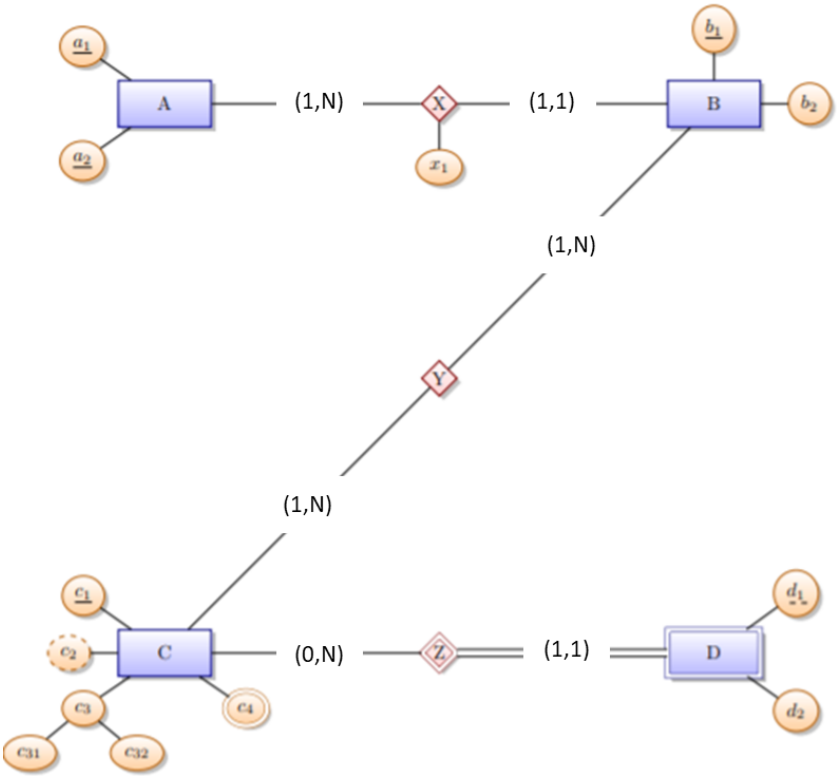
(c) (2 marks): Primary index is built on prime attribute and clustering index is built on non-prime attribute (An attribute is prime attribute if it belongs to one candidate key).

(d) (2 marks): In an extendible hashing system, each hash code is a bit string value that is associated with exactly one bucket.

(e) (2 marks): Given two relations R(a,b,c) and S (c,d,e), the results of R*S and R $\bowtie_{R.c=S.c}$ S are not equivalent.

Question 2 (24 marks): Database Design

(a) (12 marks) Take the following ER-model and translate it into a relational schema using the rules presented in class. Present the relational schema using the notation from the slides. For example, a relation R with attributes a1 and a2 where a2 is the primary key is written as R(a1, a2). You also need to specify foreign key constraints by arrows. You do not have to show intermediate results.



(b) (12 marks) Consider the relation with schema R (A, B, C, D, E, F) and the following functional dependencies (FDs): $A \rightarrow BC$, $D \rightarrow AF$

(1) (4 marks) Please write down the candidate keys of the relation, and proof it with closure of attribute.

(2) (4 marks) Is relation R in BCNF? If it is, explain why it is. If it is not, explain why not and give a decomposition of R into a collection of relations that are in BCNF.

(3) (4 marks) About 3NF and BCNF, which of the following statements are **TRUE**? (Multiple choices)

(A) Every relation in 3NF is also in BCNF

(B) A relation R is in 3NF if every non-prime attribute of R is fully functionally dependent on every key of R

(C) Every relation in BCNF is also in 3NF

(D) No relation can be in both BCNF and 3NF

(E) If a 3NF table has only one candidate key, then this table is also in BCNF.

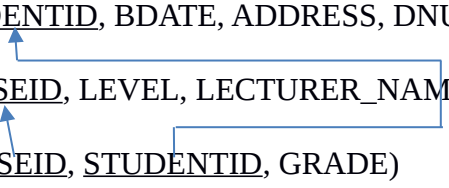
Question 3 (30 marks): Querying

(a) (6 marks) Consider the following relations containing students and courses information:

STUDENT(SNAME, STUDENTID, BDATE, ADDRESS, DNUM)

COURSE(CNAME, COURSEID, LEVEL, LECTURER_NAME)

COURSE_TAKING(COURSEID, STUDENTID, GRADE)



Note the primary key of each relation is underlined, and the foreign key reference is indicated by arrows. Translate the relation definition into create table statements and include all necessary primary key and foreign key constraints.

(b) (12 marks) Based on the schema in (a), compose SQL statements for the following queries.

(1) (4 marks) Display the course id, name and the number of students taking each course.

(2) (4 marks) Display the student id, student name and the GPA of each student who score lower than 60. Assume that all courses carry the same credit weight and the possible grades of each course are in the range of [0,100].

(3) (4 marks) Display the name of the student, if he/she has the same name as his/her lecturer.

(c) (12 marks) Based on the schema in (a), write down the relational algebra expressions for the following queries.

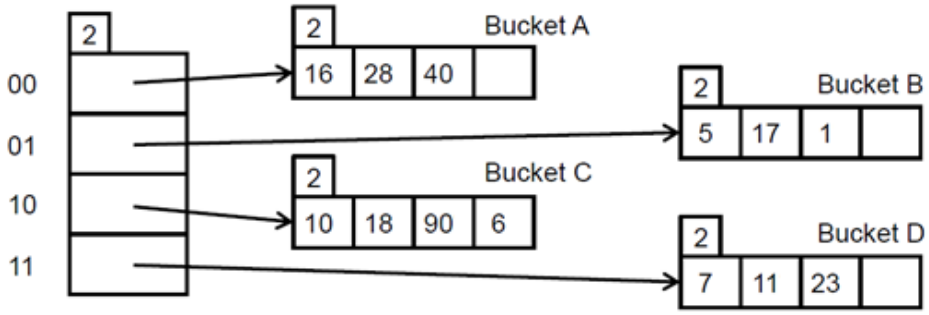
(1) (4 marks) Display id of the course with no enrolment.

(2) (4 marks) Display id of the student who takes all courses ‘Mary’ takes.

(3) (4 marks) Display the name of the student, if the student gets the highest grade on some course he/she takes.

Question 4 (16 marks):

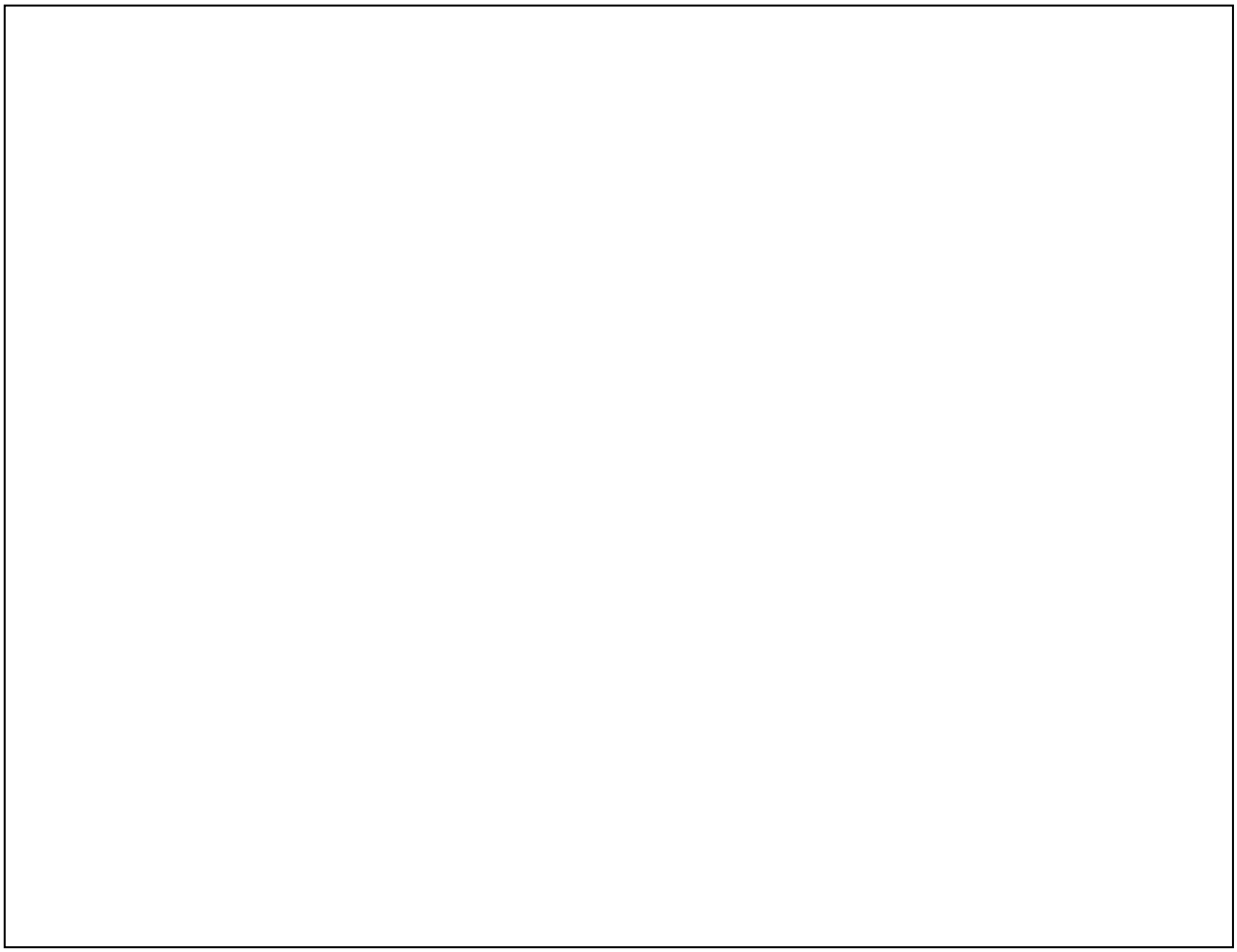
(a) (8 marks)Consider the following extendible hash.



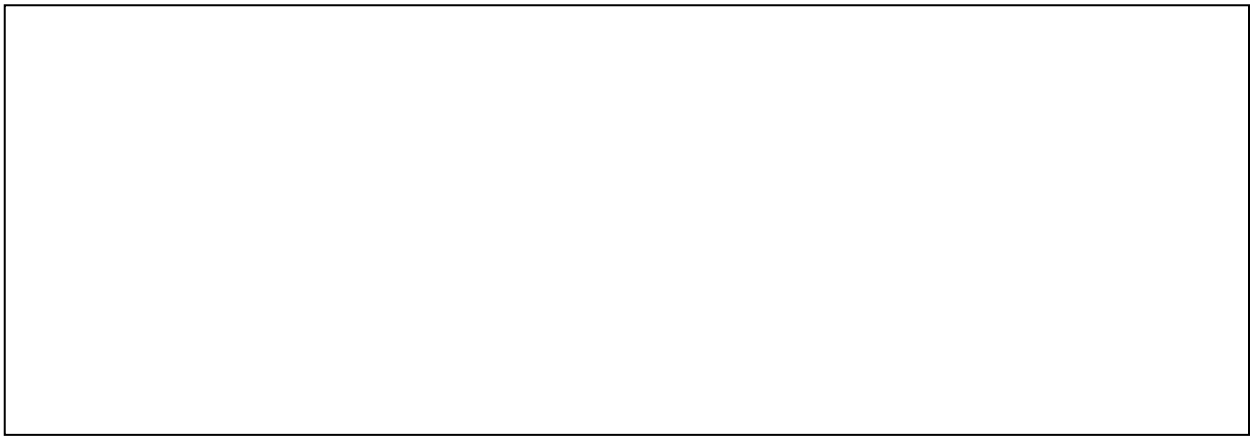
(1) (4 marks) Draw the index after the following keys have been inserted: 4, 21, and 25.

Hints: decimal-binary conversion table

Decimal	Binary
1	1
4	100
5	101
7	111
10	1010
11	1011
16	10000
17	10001
18	10010
21	10101
23	10111
25	11001
28	11100
40	101000
90	1011010



(2) (4 marks) Having inserted the above three keys in Question4(1), what is the minimum number of delete operations needed for the global depth to decrease? And which keys should be deleted to achieve that? (Hint: A bucket is merged with its split image if and



only if it becomes empty.)

(b) (8 marks)Consider a relational database with two table schemes:

Course (c-name, room, instructor)

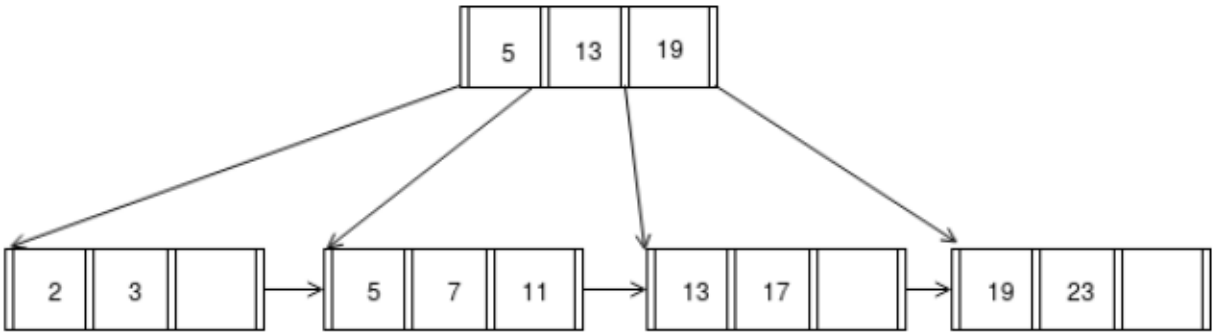
Enrollment (student-name, c-name, grade)

(1) (4 marks) To facilitate as efficiently as possible such a query like "given a course, find out who are the students taking the course", what file structure will you recommend to use, and why?

(2) (4marks) To support as efficiently as possible such a query like "given an instructor, find out who are the students taught by him/her", what kind of indexing will you recommend to use and how to use it?

Question 5 (20 marks): Indexing

(a) (10 marks) Consider the following B^+ tree with order d=4 and height h=2.



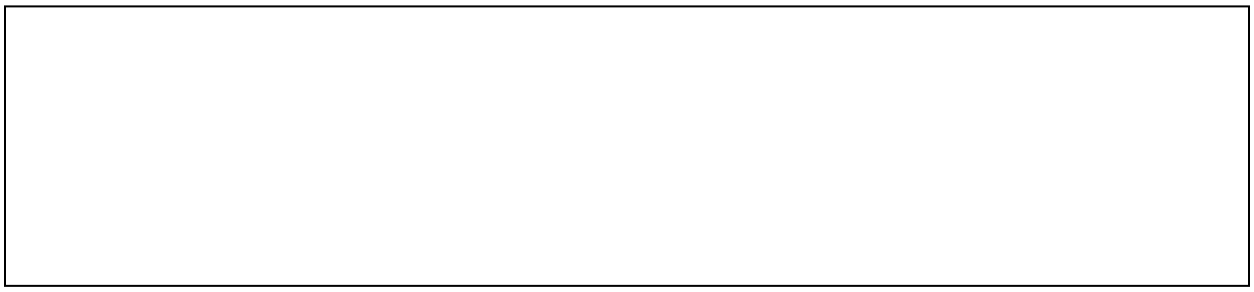
When answering the following questions, please follow the assumptions:
• A left pointer in an internal node guides towards keys < than its corresponding key, while a right pointer guides towards keys

• A leaf node underflows when the number of **keys** goes bellow $\lceil \frac{d-1}{2} \rceil$.

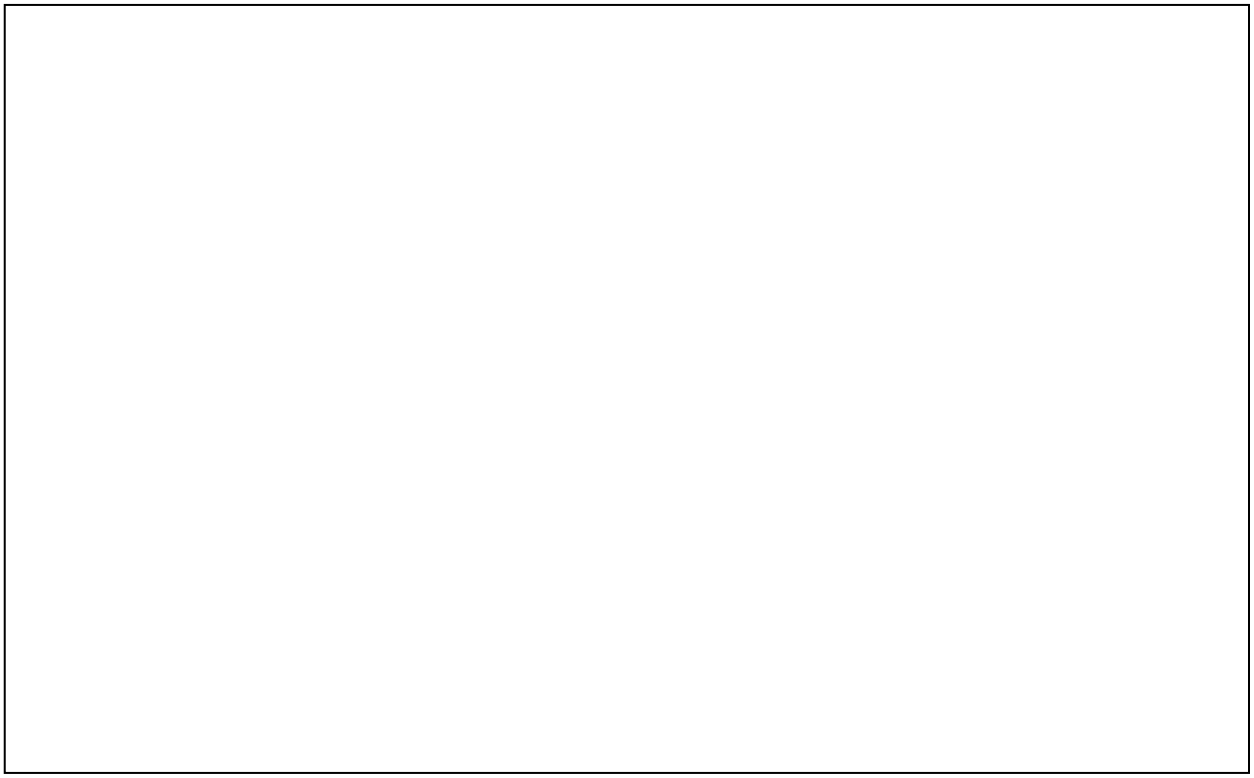
• An internal node underflows when the number of **pointers** goes below $\lceil \frac{d}{2} \rceil$.

(1) (4 marks) Insert 10 into the B+tree. Draw the resulting tree.

(2) (2 marks) After (1), how many pointers (parent-to-child and sibling-to-sibling) do you chase to find all keys between 5 and 15?



(3) (4 marks) Then delete 23. Draw the resulting tree?



(c) (10 marks) Consider a disk with block size B=256 bytes. A block pointer is P=7 bytes long, and a record pointer is P R =7 bytes long. A file has r=30,000 EMPLOYEE records of fixed-length. Each record has the following fields: NAME (30 bytes), SSN (9 bytes), DEPARTMENTCODE (9 bytes), ADDRESS (40 bytes), PHONE (10 bytes), BIRTHDATE (8 bytes), SEX (1 byte), JOBCODE (4 bytes), SALARY (8 bytes, real number). An additional byte is used as a deletion marker. Suppose the file is stored with an unspanned organization and file is ordered by the key field SSN.

(1) (2 marks) Calculate the record size R in bytes and the blocking factor bfr.

(2) (2 marks) If we want to construct a single-level primary index on SSN, calculate the index blocking factor bfr_i;

(3) (2 marks) Based on (1)&(2), calculate the number of index entries and the number of index blocks.

(4) (4 marks) If we want to construct a multi-level index on SSN, calculate the number of levels needed and the total number of blocks required.