Name: XIANG Yutan Student ID: 23096512

THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF COMPUTING

EXAMINATION

Course : Broad Discipline of COMP-61431, BSc Internet & MT-42477,
BSc Investment Science & Finance Analytics-63426,
BSc Data Science & Analytics-63428-SYD

Subject : COMP2411 Database Systems

Group : 1011, 1012, 141, 182

Session : 2022 / 2023 Semester I

Date : 13 December 2022 Time : 12:30 - 14:30

Time Allowed: 2 Hours Subject Lecturer: Prof. LI Qing / Dr WEI Junqiu

This question paper has 9 pages (cover included).

Instructions to Candidates:

- No late submission and re-submission.
- This is a close-book closed-notes exam.
- Answer ALL questions in the space provided; use backside space if needed.
- Write down your name and student ID on each page on this question paper.
- All the answers must be written in English.
- Please do not cooperate with anyone. If any plagiarism cases are detected, the Departmental Learning and Teaching Committee may investigate these cases. All the students involved in these cases will receive 0% for this exam.

Do not turn this page until you are told to do so!

Name: _____

Student ID: _____

Problem one: Basic Concepts (25 points)

1) Judge whether each of the following statements is TRUE or FALSE and give your justification. (15 points)

(a) By using the heuristic optimization, we can always transform a query into the best query regarding the query execution time.

F. We don't know exact time unless we try, so we can only choose cost less one

(b) When designing tables, if lossless join is guaranteed, then the higher the normal form is, the better performance the future queries will have on the database.

F.. Not exactly Higner form only mean more standard and splic in more ~ small tables.

(c) We cannot associate attributes to relationships.

F. ↳ close related.

(d) Primary index is a dense index if each data file block only contains one record.

T. number of index = number of block = num record

(e) We can insert rows into a "view" no matter how many tables this view is based on.

F. can't change view.

2) Use one sentence to explain what aspect of database design the following topics cover (or what you know about database design after learning the following topics). (10 points)

(a) Normalization (2 points)

1NF (Atomity in domain) → 2NF (No partial order) → 3NF (no transitive) → BCNF(4NF) (all left is primary key)

(b) SQL (2 points)

(c) File Organization and Index (2 points)

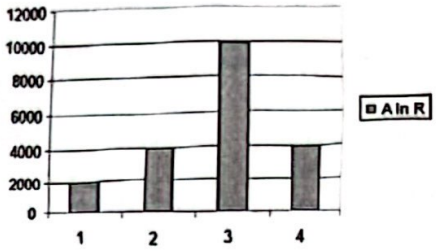
(d) Query Optimization (2 points)

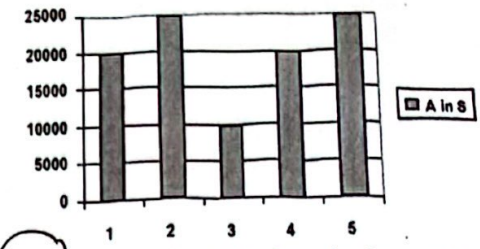
(e) Transactions and Concurrency Control (2 points)

Student ID: _____ Name: _____

Problem Two: Query Optimization (25 *points*)

1) You are given two tables: R(A, B, C) and S(A, B, Y). R contains 20,000 records, and S contains 100,000 records. The possible values of attribute A in R are: {1,2,3,4}, whereas the possible values of A in S are {1,2,3,4,5}. The following histograms present statistical information about the occurrences of values for A in R and S, e.g., there are 2,000 records with A=1 in R and 20,000 records with A=1 in S. (12 points)





(a) How many records (i.e., cardinality of the output size) are there in the result of the query (R *A S) ? (3 points)

$10^2 \cdot 10^3 (2 \cdot 20 + 4 \cdot 25 + 10 \cdot 10 + 4 \cdot 20) = 3.2 \cdot 10^8$

(b) Given that R.B is a NOT NULL foreign key referencing S.B, how many records are in the result of the query (R *B S) ? (3 points)

20000

(c) How many records are there in the result of $((\sigma_{A=1}R)$ *B S)? (3 points)

2000

(d) How many records are expected in the result of $((\sigma_{A=1}R)$ *B $(\sigma_{A=3}S))$? (3 points)

$2000 \cdot \frac{10^4}{10^5} = 200$

Student ID: _____

2) Consider the following two tables:

Sailors(Sid, Name, Rating, Age)
Reserves(Sid, Bid, Date)

Each attribute (and pointer whenever applicable) is 20 bytes. Each block is 1000 bytes. There are 10,000 sailors, 40,000 reservations and a memory of M=1000 blocks.

Assume the query: Find the names of sailors who have reservations

–SELECT Name
–FROM Sailors, Reserves
–WHERE Sailors.Sid=Reserves.Sid

(13 Points)

(a) How many bytes does each Sailors record have? How many bytes does each Reserves record have? (4 Points)

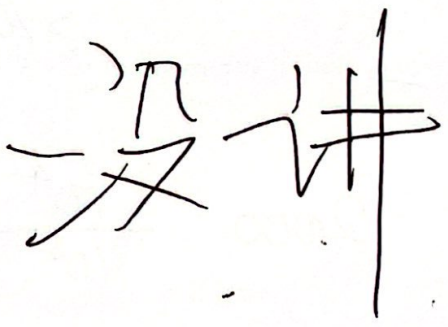
$20 \cdot 4 = 80$ $20 \cdot 3 = 60$

(b) How many blocks does Sailors contain? How many blocks does Reserve contain? (4 Points)

$bfr_S = \lfloor \frac{1000}{80} \rfloor = 12.$ $bfr_R = \lfloor \frac{1000}{60} \rfloor = 16$

$b_S = \lceil \frac{10000}{12} \rceil = 834$ $b_R = \lceil \frac{40000}{16} \rceil = 2500$

(c) Consider the two-step Hash-Join that we learned in the lecture. Assume that the records of files Sailors and Reserves are hashed using the same hashing function on the join attributes Sid of Sailors and Sid of Reserves (the number of buckets is 10). Consider two plans for this Hash-Join: (i) We hash Sailors first, and then Reserves, and (ii) We hash Reserves first, and then Sailors. Which plan has smaller number of block reads? Please also describe in details the two steps involved in Hash-Join in the plan with smaller number of block reads. (5 Points)



Student ID: _____ Name: _____

Problem Three: File Organization and Indexing (25 points)

1) Explain why the allocation of records to disk blocks is an important issue in database system performance. **(5 points)**

memory is limit → store size small √.
time is limit. use effice index can speed up.. ooooo.

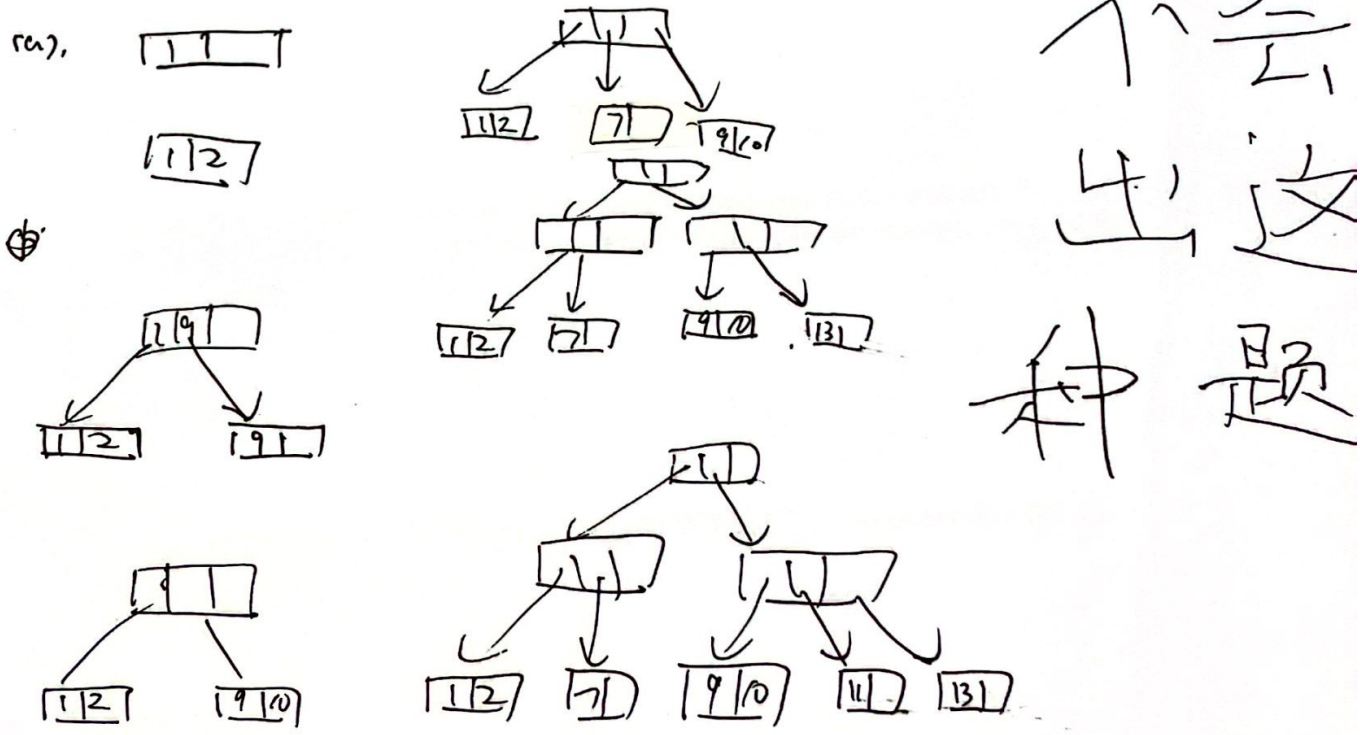
2) For hashing files, if a disk block becomes empty as a result of deletions, for what purposes should the block be reused? Why? **(6 points)**

3) Insert the following numbers one by one in sequence into a B+ tree with order 3 and leaf order 2:

1, 2, 9, 10, 7, 13, 11

(a) Show the picture after each insertion.
(b) Show the picture after each of the following deletions:
9, 11, 13

(14 points)



Name: _____

Problem Four: Transaction Processing and Concurrency Control (25 points)

CONSIDER TWO TRANSACTIONS T1 AND T2. PLEASE NOTE THAT WE USE THE NOTATIONS R1(X) AND W1(X) TO DENOTE THE READ AND WRITE OPERATIONS OF T1 ON A DATA ITEM X. SIMILARLY, WE USE THE NOTATIONS R2(X) AND W2(X) TO DENOTE THE READ AND WRITE OPERATIONS OF T2 ON A DATA ITEM X.

1) Is the schedule W2(B) R1(A) W1(A) R2(A) C1 C2 recoverable and cascadeless? (C1 and C2 indicate the commit statements of T1 and T2, respectively). Please elaborate the reasons of the answers. (4 Points)

没明白 啥?

2) Is the schedule W2(B) R1(A) W1(A) R2(A) C1 C2 serializable? Please show the precedence graph of the schedule. Please give its equivalent serial schedule if it is serializable. Please provide the reason with the precedence graph if it is not serializable. (4 Points)

Yes

T1, T2

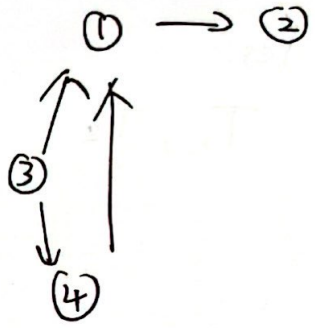
3) If you are allowed to move the commit statements in the schedule W2(B) R1(A) W1(A) R2(A) C1 C2, how would you change them so that the schedule will become a cascadeless schedule? (4 Points)

4) Is the schedule R2(A) R1(A) W1(A) W2(B) C2 C1 recoverable and cascadeless? (4 Points)

Student ID: _____ Name: _____

5) Is the following schedule conflict serializable? What is the equivalent serial schedule? Please show the precedence graph of the schedule. (4 Points)

T1	T2	T3	T4
Read(X)			
Write(X)			
	Read(X)		
		Read(Y)	
		Write(Y)	
	Write(X)		
			Read(Y)
Write(Y)			



Yes.

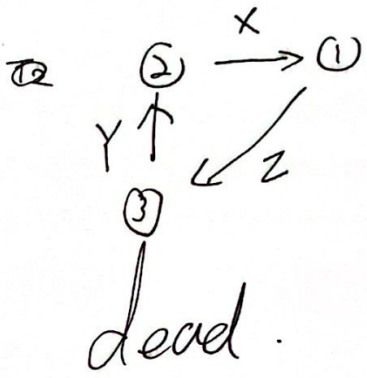
T3 T4 T1 T2.

6) Consider the following schedule assuming the timestamps 1, 2, 3 for transactions T1, T2, T3 respectively. Show the wait-for-graph of this schedule and determine if there is a deadlock. Note that we use the notation R() and W() to denote the read_lock and write_lock, respectively. (5 Points)

ent ID: _____

Name: _____

T_1 TS=1	T_2 TS=2	T_3 TS=3
R (X)		
	R(Y)	
	W(Y)	
		W(Z)
W(X)		
	R(X)	
	W(X)	
		R(Y)
		W(Y)
W(Z)		



-END-