考试方式：开卷

考试时长：120 分钟　　　　　　院　系：软件工程

19 软件工程

警示 《中山大学授予学士学位工作细则》第八条："考试作弊者，不授予学士学位。"

------------以下为试题区域，共 6 道大题，总分 100 分，考生请在答题纸上作答------------

Question 1. (10 marks) For each of the following statements, determine its truth value (TRUE or FALSE):

a) (2 marks) 1NF allows storing lists as an attribute value.

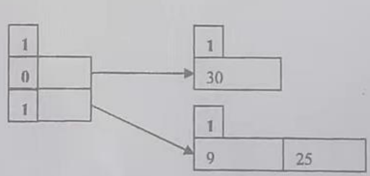
b) (2 marks) In relational query processing, physical query plans are translated into logical ones by the query optimizer. X PaBer

c) (2 marks) With 2PL, once a transaction releases a lock, it can acquire a new lock. X

d) (2 marks) Sort-Merge Join can be better than Index Nested Loops Join. ✓

e) (2 marks) All natural joins are also theta joins.
θ

Question 2. (20 marks) You are given an initial extendible hash structure with three keys already inserted as below. You should follow the convention that binary hash indices start from the **least significant bit**. Draw 4 extendable hash structures (**each 5 marks**) corresponding for each insertion of the following search key values: 23, 15, 52, 37. Assume each bucket can hold two keys and the search key values arrive in the given order (i.e. 23 being the first coming key and 37 being the last one).



23

23÷7

15÷8=7

9÷4 = 1

25÷4 = 1

23÷4=3

37÷8

432

5

37　37÷

37÷8=5　4936

Question 3. (16 marks) Consider a relation R=(Title, Price, Shop, City) where Title is the name of a book, Price is the book price, Shop is the name of a book shop and city is the city where the book shop is located. Given the following constraints: The book title determines the book price. The shop name determines the city where the book shop is located. If two different shops are located in the same city, then the two shops can not sell the same book. A shop can sell many books.

key · 10000+

a) (6 marks) Write the set of functional dependencies implied by the above constraints.

b) (4 marks) Identify the candidate key(s), and is R a 3NF or BCNF relation?

c) (6 marks) If R is not BCNF, decompose it to BCNF.

B=102

101 x100

shop, city → book

10000 ≤ (B-1)(B-2)

Question 4. (20 marks) Consider joining two relation R and S on a single joining attribute. Both relations are stored in pages where each page has a size of 8KB. Note that in the join processing, we ignore the cost of writing the final join answer out to disk. ✓

a) (10 marks) Suppose R has 1000 pages, and S has 500 pages. Assume we have 400KB of memory buffer available for joining R and S. Using Block Nested Loops Join, with S as the outer relation, show the total number of I/O's of the join.

b) (10 marks) Suppose R has 10000 pages now and S still has 500 pages, and we can uniformly partition both R and S on the joining attribute. Assume we want to store partitions of R in hash tables in the memory buffer in the probing phase of hash-join. Show the least KB of memory buffer that we need to complete the Hash-Join without recursive partitioning of R. And in this case, 1) show the total number of I/O's of the Hash-Join (5 marks); 2) show the total number of I/O's of the Sort-Merge Join (5 marks).

Question 5. (24 marks) Consider the following schema for sport teams:

Player(pid, name, age, salary, rating) for players, where pid is the key

Team(tid, name) for teams, where tid is the key

MemberOf(pid, tid) -- pid is foreign key to Player, and tid is foreign key to Team.

a) (6 marks) Write a SQL query that returns the names of players who have a rating of 7.

b) (6 marks) Write a SQL query that returns the names of players who are currently members of at least a team.

如何去重？

c) (6 marks) for the query in b), show the corresponding relation algebra expression.

d) (6 marks) Write a SQL query that returns the names of players who are currently members of multiple teams.

Question 6. (10 marks) Consider the following schedule S that consists of operations from four transactions in order: S = < T1_R(X), T1_W(X), T2_R(X), T3_R(Y), T3_W(Y), T2_W(X), T3_R(Z), T4_R(Z), T4_W(Z), T1_W(Y), T2_R(Y)>. The notation is self-explanatory. For example, T1_R(X) means that transaction T1 reads data object X.

a) (7 marks) show all the directed edges in the precedence graph of S. Explain why or why not the schedule is conflict-serializable.

b) (3 marks) If you find that S is **serializable** in a), write down all equivalent serial schedules of S.