中山大学软件学院 **2011** 级软件工程专业(2012 春季学期)

《SE-304 数据库系统》期末考试试题(A)

(考试形式： 卷 考试时间：2 小时)

警示

《中山大学授予学士学位工作细则》第六条

考试作弊不授予学士学位

方向：_____ 姓名：_____ 学号：_____

出卷：_____ 审核：_____

注意：答案一定要写在答卷中，写在本试题卷中不给分。本试卷要和答卷一起交回。

Question 1. (20 marks) Consider the schema R(A, B, C, D) with four consistent tuples.

A	B	C	D
1	1	1	2
1	2	1	3
2	2	1	4
2	1	1	5

(i) A user tries the query:

```
INSERT INTO R VALUES (2, 2, 2, 2);
```

The users' insertion operation is aborted due to a consistency violation. Consider each of the following integrity constraints *independently* and circle it if it could have caused that violation:

a) (A, B, C) is a candidate key of R
b) (A, B) is the primary key of R
c) D→B
d) C is a foreign key

Answer: b, c, d

(ii) Ignoring the constraints in part (i), is CD a super key of R? Justify your answer.

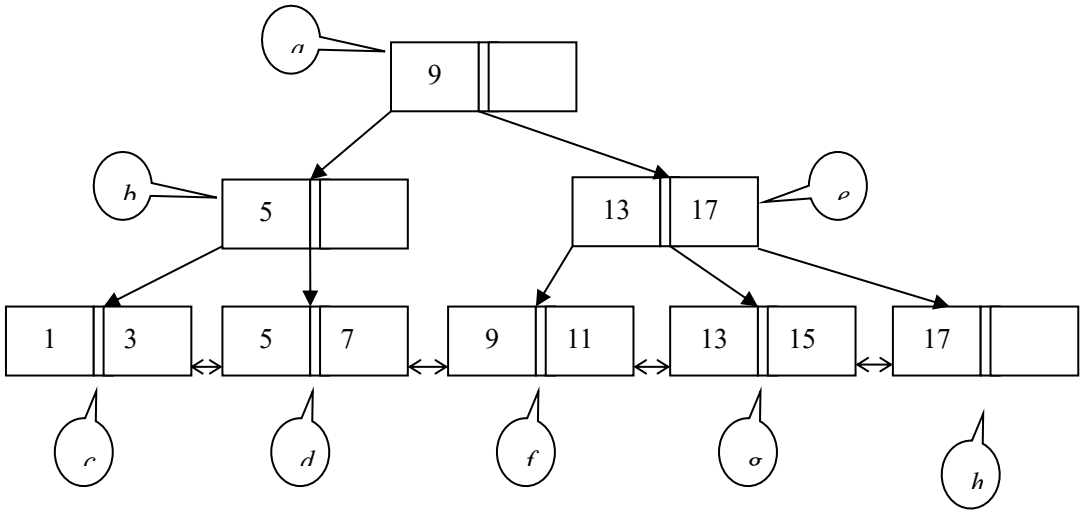
Answer: A *super key* of an entity set is a set of one or more attributes whose values uniquely determine each entity. D is a candidate key of R, and CD is a super key.

(iii) You are told that the Functional Dependencies on R are F = {AB → CD, D → B}. Write a decomposed schema for R that is in Boyce-Codd Normal Form (BCNF).

Answer:

BCNF decomposition: R1(A, C, D) and R2(B, D)

Question 2. (30 marks) Consider the following B+ tree index of order 1:



(i) Assume numbering the eight nodes in the tree from a to h as shown in the figure. List all the nodes that must be fetched to satisfy the query "Get all records with search key larger than or equal to 7 and less than 16".

Answer:

Node a, b, d, f, g, h

(ii) Assume we modify the B+ tree by adding the following keys in the following order: 20, 27, 18, 30, 19.

In the answer-boxes below, each row refers to a key being inserted in order, and each column asks if the insertion of that key results in a split of particular nodes. Assume that when splitting up a node with an odd number of entries, the left node gets one more entry than the right. Place a check mark (✓) in each box whose answer is "Yes". Blank boxes will be interpreted as "No".

Answer:

Key	Leaf Node Split?	Non-Leaf Split?	Root Split?
20	**No**	**No**	**No**
27	**Yes**	**Yes**	**No**
18	**Yes**	**No**	**No**
30	**No**	**No**	**No**
19	**Yes**	**Yes**	**Yes**

Question 3. (20 marks) Suppose there is a schema R(A, B, C), and we do not know the keys of the table.

(i) How to confirm that A is not a candidate key of R by using a SQL query?

Answer:

```
SELECT A
FROM R
GROUP BY A
HAVING COUNT(*)>1
```

If this query gives a non-empty result, then A is not a key.
If the result is empty we cannot be sure.

(ii) How to confirm that the FD A→B holds with a SQL query?

Answer:

```
SELECT A
FROM R
GROUP BY A
HAVING COUNT(DISTINCT B)>1
```

If this query gives a non-empty result, this FD does not hold.

Question 4 (20 marks) Consider the following two relations R and S. The number of records in R and S is also given.

- R(A, B, C, D): 20,000 records
- S(A, E, F): 6,000 records

Assuming the size of each memory page is 4K (i.e. 4000) bytes and the memory buffer has 50 pages. We assume all attribute values and pointers, if needed to be considered, are 10 bytes. Consider the following query:

```
SELECT    *
FROM      R, S
WHERE     R.A = S.A
```

Give estimation of the following questions. All the computations steps and their reasoning should be clearly explained.

(i) What is the size of R and S in terms of pages?

Answer:

R: 40 * 20000 / 4000 = 200 pages
S: 30 * 6000 / 4000 = 45 pages

(ii)

a) Assume using Sort-Merge Join to process the query. Estimate the cost of each step and then compute the total cost of the join.

b) Assume using Block-Nested-Loop Join to process the query. Estimate the cost of each step and then compute the total cost of the join.

c) Assume using Hash Join to process the query. We assume uniform partitioning（均匀划分）. Estimate the cost of each step and then compute the total cost of the join.

Which is better?

Answer:

a) Sort-Merge Join:

Cost:4*200+4*45+45+200=1225

b) Block-Nested-Loop Join:

Cost: R outer: 200+(200/50)*45 = 380

S outer: 45+(45/50+1)*200= 245

c) Hash Join：

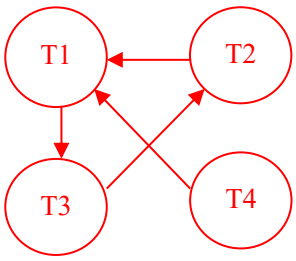
Cost: 3*(200+45) = 735

Question 5 (10 marks) Consider the schedule S that consists of four transactions as follows: S = < T2_X(A), T2_W(A), T1_S(B), T1_R(B), T2_X(B), T3_S(C), T3_R(C), T1_X(C), T4_X(B), T4_W(B), T3_S(A)>. The notation is self-explanatory. For example, T1_R(B) means that transaction T1 reads item B.

T1	T2	T3	T4
	X(A)		
	W(A)		
S(B)			
R(B)			
	X(B)		
		S(C)	
		R(C)	
X(C)			
			X(B)
			W(B)
		S(A)	

Show the waits-for graph（等待图） and determine whether there is a deadlock. How to resolve the deadlock（解除死锁） if it exists?

Answer:



There is a deadlock since a circle exists in the waits-for graph.
A deadlock is resolved by aborting a transaction that is on a cycle and releasing its locks.