**诚信应考,考试作弊将带来严重后果！**

姓名 学号 学院 专业 座位号



( 密 封 线 内 不 答 题 )

……………………………………………………密………………………………………………封………………………………………线……………………………………线………………………………………

**华南理工大学期末考试**

**《数据库系统》试卷A**

**注意事项：1. 考前请将密封线内填写清楚；**

**2. 所有答案请直接答在试卷上；**

**3．考试形式： 闭 卷；**

**4. 本试卷共 大题，满分100分， 考试时间120分钟**。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **题 号** | **Part I** | **Part II** | | | | | **总分** |
| **1** | **2** | **3** | **4** | **5** |
| **得 分** |  |  |  |  |  |  |  |
| **评卷人** |  |  |  |  |  |  |  |

**Part I [20 pts.] (1pt each) Fill in the blanks with the best answer**

1. The collection of information stored in the database at a particular moment is called an instance of the database. The overall design of the database is called the database schema.
2. A relation schema R is in third/3 normal form if for all α → *β* in F+,at least one of the following holds: α → *β* is trivial; α is a superkey for R;Each attribute A in α → *β* is contained in a candidate key for R
3. Let R be a relation schema, R1 and R2 form a decomposition of R. Decomposition is a lossless\_ if for all legal databases instances r of R, .
4. Assume relation *r* has *br* blocks and relation *s* has *bs* blocks, therefore, in the best case, only *br* + *bs* block transfers would be required for .
5. An ideal hash function is uniform and random, the former require that each bucket is assigned the same number of search-key values from the set of all possible values.
6. To generate query-evaluation plans for an expression, we have to generate logically equivalent expressions using equivalence rules.
7. Consider a B+-tree of order n, if there are K search-key values in the file, the path from the root to the leaf node is no longer than \_⎡log⎡n/2⎤K⎤\_ .
8. Hash indices are always secondary (primary or secondary) indices.
9. We assume that a relation has a B+ tree index of height 5, each disk block contains 4 tuples of the relation, and there are 10 tuples satisfying a selection operation. To process the operation, database management system have to access disk at least 15 times in the worst case.
10. A transaction has the following properties: atomicity, consistency, isolation and durability.
11. The immediatedatabase modification scheme allows database modification to be output to the database while the transaction is still in active state.
12. In the process of deadlock prevention, wound-wait scheme uses a preemptive strategy, that is, older (older/younger) transaction wounds (forces rollback) of younger (older/younger) transaction instead of waiting for it.
13. In deferred database modification scheme, redo operation is the only operation used in the recovery procedure.
14. The system is in a deadlock state if and only if the wait-for graph has a cycle.
15. Cascading rollbacks can be avoided by applying strict/rigorous two-phase locking protocol to transactions in currency control.

Answer:

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
5. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
6. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
7. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
8. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
9. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
10. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
11. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
12. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
13. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Part II [80 pts.] Answer the following questions**

1. [14points]Database design I: Consider the following conditions
   * 1. The STUDENT may be taught by one and only one teacher. The TEACHER may be instructor of one or more STUDENTS.
     2. The TEACHER may be responsible for one and only one CLASS. The CLASS may be the responsibility of one and only one TEACHER.
     3. The CLASS may be made of one or more STUDENTS. The STUDENT must be a member of one and only one CLASS.
     4. The CLASS must have one and only one ROOM. The ROOM may belong to one or more CLASS.

Notes: Assume entity CLASS has the following attributes: CID and CNAME, entity ROOM has the following attributes: RID and LOCATION, entity STUDENT has the following attributes: SID, LASTNAME, and FIRSTNAME, entity TEACHER has the following attributes: TID, TEACHERNAME, and TITLE.

1. [6points]Construct an ER diagram showing these relationships.
2. [4points] Construct appropriate relation schemas for the relationships in your ER diagram.

S-T(SID,TID)

1. [4points] Create an index *std\_index* on the *Student* relation with *SID* as the search\_key.

Create Index std\_index on STUDENT(SID)

1. [6points] In database design, how to represent relationship sets as relational schemas?

A many-to-many relationship set :



Many-to-one /one-to-many relationship sets:



1. [10points]: Let *R* =(*A,B, C, D, E, F* ) be a relation withfunctional dependency *F* ={*A* → *CB, E* → *F A* } and key = *{E, D*}.
   * 1. Why we need BCNF ? [4pts]

It eliminates all redundancy that can be discovered based on functional dependencies.

* + 1. Please decompose this relation into BCNF [6pts]

*1. A* → *CB* 违反BCNF定义

*R*1 = (*A,B, C* )

*R*2 = (*A, D, E, F* )

*2. R*2中*E* → *F A*违反BCNF定义,对*R*2分解

*R*3 = (*A, E, F* )

*R*4 = (*D, E* )

Final decomposition: *R*1*, R*3*, R*4

1. [30points] BOOK (Bookid,Title,Publishername)

BOOK\_AUTHORS(Bookid,Authorname)

PUBLISHER(Publishername, Address, Phone)

BOOK\_COPIES(Bookid,Branchid,No\_Of\_Copies)

LIBRARY\_BRANCH(Branchid,Brachname,Address)

BOOK\_LOANS(Bookid, Branchid, Cardno,DateOut,Duedate)

BORROWER(Cardno,Name, Address,Phone)

1. [4points]Write appropriate SQL DDL statements for declaring the BOOK\_AUTHORS relation.
2. [6points]Give an expressions in relational algebra to express the following queries:

Q1:Retrieve the names of all borrowers who do not have any books checked out.

Answer: (Note: We will use S for SELECT, P for PROJECT, \* for NATURAL JOIN, - for SET DIFFERENCE, F for AGGREGATE FUNCTION)

NO\_CHECKOUT\_B <-- P CardNo (BORROWER) - P CardNo (BOOK\_LOANS)

RESULT <-- P Name (BORROWER \* NO\_CHECKOUT\_B)

Q2:For each book that is loaned out from the "Sharpstown" branch and whose DueDate is today, retrieve the book title, the borrower's name, and the borrower's address.

S <-- P BranchId ( S BranchName='Sharpstown' (LIBRARY-BRANCH) )

B\_FROM\_S <-- P BookId,CardNo ( ( S DueDate='today' (BOOKLOANS) ) \* S )

RESULT <-- P Title,Name,Address ( BOOK \* BORROWER \* B\_FROM\_S )

1. [16points]Give an expressions in SQL to express the following queries:

Q1:How many copies of the book titled The Lost Tribe are owned by the library branch whose name is "Sharpstown"?

SELECT NoOfCopies

FROM ( (BOOK NATURAL JOIN BOOK\_COPIES ) NATURAL JOIN

LIBRARY\_BRANCH )

WHERE Title='The Lost Tribe' AND BranchName='Sharpstown'

Q2:For each library branch, retrieve the branch name and the total number of books loaned out from that branch.

SELECT L.BranchName, COUNT(\*)

FROM BOOK\_COPIES B, LIBRARY\_BRANCH L

WHERE B.BranchId = L.BranchId

GROUP BY L.BranchName

Q3: Retrieve the names, addresses, and number of books checked out for all borrowers who have more than five books checked out.

SELECT B.CardNo, B.Name, B.Address, COUNT(\*)

FROM BORROWER B, BOOK\_LOANS L

WHERE B.CardNo = L.CardNo

GROUP BY B.CardNo

HAVING COUNT(\*) > 5

Q4: For each book authored (or co-authored) by "Stephen King", retrieve the title and the number of copies owned by the library branch whose name is "Central".

SELECT TItle, NoOfCopies

FROM ( ( (BOOK\_AUTHORS NATURAL JOIN BOOK) NATURAL JOIN

BOOK\_COPIES)

NATURAL JOIN LIBRARY\_BRANCH)

WHERE Author\_Name = 'Stephen King' and BranchName = 'Central'

d.[4points] Record the fact that the manager didn't maintain information about the book named “T&G”, i.e. remove information about “T&G”.

1. [20pts]Query Processing, Optimization and Transaction
   1. [5points] Please describe the implementation process of selection operation , where r is a relation, A is an attribute and is not a candidate key, r has a primary index on A. If there are n matching records, the B+ tree index is of height h, and each disk block contains at most d records, please analyze the overhead in the best case.

Process:

1. The algorithm walks from the root of the B+ tree to the leaf containing valuec.

2. According the address given by the leaf, the algorithm retrieves the n matching records.

Overhead:

1. It accesses disk h times. in each time, it needs a seek and a block transfer.

2. In the best case, these n records are distributed on n/d disk blocks. Thus it seeks disk 1 time and transfer block n/d times.

* 1. [4points] Describe the process of Indexed nested-loop join

.

* 1. [5points] Please describe the two-phase locking protocol and prove that it ensures conflict-serializable schedules and does not ensure freedom from deadlocks.
  2. [6points] Below we show some log of a DBMS, please describe the recovery procedure using immediate database modification.

<*T*0**start**> <*T*0**start**> <*T*0**start**>

<*T0,* A, 1000, 950> <*T0,* A, 1000, 950> <*T0,* A, 1000, 950>

*<T*o*,* B, 2000, 2050> *<T*o*,* B, 2000, 2050> *<T*o*,* B, 2000, 2050>

<*T*0 **commit**> <*T*0 **commit**>

<*T*1 **start**> <*T*1 **start**>

<*T*1, C, 700, 600> <*T*1, C, 700, 600>

<*T*1 **commit**>

(a) (b) (c)