浙江大学 2019－2020 学年 春夏 学期

《数据库系统》课程期末考试试卷

课程号：21121350， 开课学院：计算机学院

考试试卷：√A 卷、B 卷（请在选定项上打√）

考试形式：√闭、开卷（请在选定项上打√），允许带一张 A4 纸笔记入场

考试日期：2020 年 9 月 5 日，考试时间：120 分钟

诚信考试，沉着应考，杜绝违纪。

考生姓名：______________学号：______________所属院系：____________

题序	一	二	三	四	五	六	七	八	总 分
得分									
评卷人									

Problem 1：Relational Model and SQL（18 points)

A software development company develops software projects for different clients. It has the following relational schemas for its internal management system:

client (<u>cId</u>, cName, cCity)
project (<u>pId</u>, pName, cId, startTime, endTime, budget, paid)
employee (<u>eId</u>, eName, eAddress, eSalary, eBonus)
participate (<u>pId, eId</u>, role)

The underlined are primary keys. “cId” in “project”, “pId” and “eId” in “participate” are foreign keys. “paid” is the cost that the client has paid for the project. One employee can participate in different projects with a specific role for each project. Only three different roles are permitted: “project manager”, “developer”, and “tester”. Based on the above schemas, answer the following questions:

(1) Write a relational algebra expression to find the project names that the client “X Bank” has, where “X Bank” is the client name. (3 points)

(2) Write a SQL statement to create table participate with all the necessary constraints. (5 points)

(3) Write a SQL statement to find the employee names who participate in different projects with all the three different roles. (3 points)

(4) Write a SQL statement to find the employee names who have the maximum salary in project “p1102”, where “p1102” is the project id. (3 points)

(5) If there are more than three participating roles and the software company requires to maintain a permitted role list in the application, what will be your suggestion to modify the schemas? (4 points)

Answers of Problem 1：

Problem 2：E-R Model （11 points)

Suppose you want to design a simple video sharing web site. **Users** can upload and watch **videos**. There are **channels** for videos. A video may belong to several channels. Users can subscribe to different channels. Channels have hierarchy structure, e.g., “chicken eating” channel is the sub channel of “game” channel. For better user recommendation, one channel may have relationships with any other channels, e.g., “travel” channel has relationship with “photography” channel, although “travel” is not a super or sub channel of “photography”.

Please draw an E-R diagram for the database design of the web site.

Answers of Problem 2：

Problem 3: Relational Formalization (12 points)

For relation schema R (A,B,C,D,H,I) with functional dependencies set F = {B->C, AC->D, BC->H}. Answer the following questions:
1) Find all the candidate keys; (3 points)
2) Find the canonical cover Fc; (3 points)
3) If R is not in BCNF, decompose it into BCNF schemas. (4 points) Is this decomposition dependency preserving? (2 points)

Answers of Problem 3：

Problem 4: XML(9 points, 3 points per part)

The following is a simplified DTD for the software company in **problem 1**:

```
    <!DOCTYPE      software_company[
        <!ELEMENT     software_company( client*,employee+)>
        <!ELEMENT     client (cname, ccity, project*)>
        <!ATTLIST     client cid ID #REQUIRED>
        <!ELEMENT     project (pname,paid)>
        <!ATTLIST     project pid ID #REQUIRED>
        <!ELEMENT     employee (ename, esalary, participate*)>
        <!ATTLIST     employee eid ID #REQUIRED>
        <!ELEMENT     participate (role)>
        <!ATTLIST     participate pid IDREF #REQUIRED>
        <!ELEMENT     cname (#PCDATA)>
        <!ELEMENT     ccity (#PCDATA)>
        <!ELEMENT     pname (#PCDATA)>
        <!ELEMENT     paid (#PCDATA)>
        <!ELEMENT     ename (#PCDATA)>
        <!ELEMENT     esalary(#PCDATA)>
        <!ELEMENT     role(#PCDATA)>
    ]>
```

Please answer the following questions:

(1) Give an XPath expression to find the project names that the client “X Bank” has paid for more than 50000 Yuans each. “X Bank” is a client name.

(2) Give an XPath expression to find the project names that employee “John” participates in as project manager. “John” is an employee name.

(3) Give an XQuery expression to find the project names from clients in “Shanghai” and employee “John” participates in. “Shanghai” is a city. “John” is an employee name.

Answers of Problem 4：

Problem 5: B+ -Tree (10 points)

Table employee in **Problem 1** is stored sequentially on eId. To build a B+-tree for table employee on column eName, buckets are used between index entries in leaf node and data records in data file. One bucket is a collection of pointers that point to data records with same employee name. One bucket is stored in one block except it exceeds the block size. If there are 13 employees initially as follows：

Wu, Mozart, Einstein, Mark, Susan, Gold, Katz, Singh, Crick, Mark, Brandt, Kim, Gold

(1) Suppose the max number of pointers in a B+-tree node is 4 (n = 4), draw a B+-tree with buckets based on the above employee names. (5 points)

(2) Suppose there are 20000 employees, within which there are 18000 distinct employee names. B+-tree node size is 4096 (n will be far bigger than 4). Employee name length is 32, pointer length in B+-tree node is 4. There are 5 employees with the same name “Mark”. To find all the information of employees named “Mark” using B+-tree with buckets, in the worst case, how many blocks will be read? (5 points)

Answers of Problem 5：

Problem 6: Query Processing (15 points)

For the relational schemas in **problem 1**, there are following information:

- Record numbers: n_{client}=500 , $n_{project}$=2000, $n_{employee}$=30000, $n_{participate}$=40000
- Records in a block: f_{client}=80, $f_{project}$=50, $f_{employee}$=40, $f_{participate}$=100
- Distinct values: V(cCity, client)=100, V(cId, project)=400, V(eId, participate)=28000, V(pId, participate)=2000
- block size is 4K bytes.
- Numbers of clients are assumed to be the same for different cities.

(1) Estimate the size (i.e. number of records) returned by the following SQL statements:

a. **select** cId **from** client **where** cCity = "Shanghai"; (3 points)

b. **select** e.eId, e.eName, pro.pId, pro.pName, par.role
from client as c, project as pro, employee as e, participate as par
where c.cId = pro.cId and pro.pId = par.pId and e.eId = par.eId and c.cCity="Shanghai"; (3 points)

(2) Estimate the number of blocks for employee and participate respectively. (4 points)

(3) Suppose there are 50 buffer blocks to use, estimate the cost of "employee natural join participate" using hash join. (5 points)

Answers of Problem 6：

Problem 7: Concurrency Control (11 points)

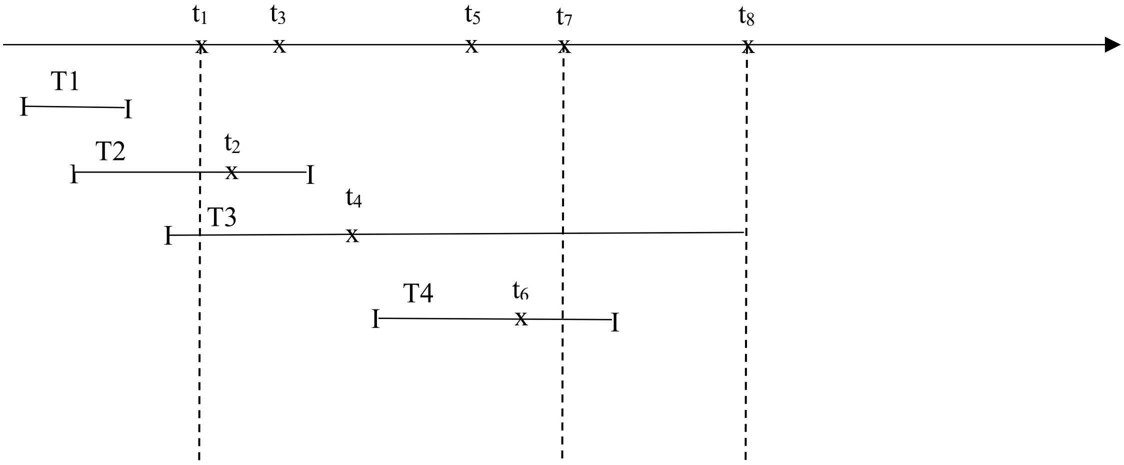
For a database with problem 1, suppose there are three transactions executing concurrently:
T1: select sum(eSalary) from employee;
T2: update employee set eSalary = eSalary + 800 where eId = “190012”;
T3: update employee set eSalary = eSalary + 500 where eId = “190012”;
update employee set eSalary = eSalary + 1000 where eId = “181020”;
The total salary of all the employees is 250000 Yuans before the execution of these three transactions. The database management system has implemented strict two-phase locking for concurrent control with record level locks only. Please answer the following questions:

(1) If all the transactions commit successfully, what are the possible outputs for T1? (4 points)
(2) Is there any possibility of deadlock? If there is, can you illustrate one transaction schedule that will cause the deadlock. (5 points)
(3) T4 inserts some new employees into table employee. If T4 executes concurrently with T1, T2 and T3, can the above-mentioned DBMS ensure serializable schedule? (2 points)

Answers of Problem 7：

Problem 8: Recovery (14 points)

A fuzzy checkpoint finds a list of modified buffer blocks (LBlocks) and a list of active transactions (LTransactions); writes a checkpoint log record with LTransactions in it; works with other transactions concurrently to output buffer blocks in LBlocks to the disk; and saves the pointer of the checkpoint log record as last-checkpoint information to a safe place on disk. Below is the timeline of a fuzzy checkpoint and related transactions. The arrow line is the time line. t_1 to t_8 are points of time ($t_1<t_2 \ldots <t_8$). T1 to T4 are transactions. "I" represents the beginning or end of a transaction. Commit of a transaction does not mean that all the data the transaction modified should be output to the disk.



Here is what happened at each of the time point:
t_1: Fuzzy checkpoint began. There were two modified buffer blocks at that time: B1 and B2; t_2: T2 updated data in B1; t_3: Checkpoint output B1 to the disk; t_4: T3 updated data in B3; t_5: Checkpoint output B2 to the disk; t_6: T4 updated data in B2; t_7: Fuzzy checkpoint finished; t_8: System failure.

Suppose there were no other DBMS components that output log buffer or buffer blocks to the disk in the meantime. Please answer the following questions:

(1) What did the fuzzy checkpoint do to log buffer, disk log file, buffer blocks, and disk data blocks at times of t_1, t_3 and t_7 respectively? (4 points)
(2) What did T2 do to log buffer, disk log file, buffer blocks, and disk data blocks at time of t_2? (2 points)
(3) Which transactions are to be redone and which are to be undone during recovery? (4 points)
(4) Please mark on the diagram the beginning point of redo with "@" and the end point of undo with "#" during recovery. (4 points)

Answers of Problem 8: