

诚信应考,考试作弊将带来严重后果!

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. 所有答案请答在答题纸上;
3. 考试形式: 闭卷;
4. 本试卷共 两 大题, 满分 100 分, 考试时间 120 分钟。

题 号	一	二	总分
得 分			
评卷人			

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

- The collection of information stored in the database at a particular moment is called an instance of the database. The overall design of the data base is called the database schema. P32
- A relation schema R is in 3th normal form if for all $\alpha \rightarrow \beta$ in F^+ , at least one of the following holds: $\alpha \rightarrow \beta$ is transitive is a super key for R ; Each attribute A in $\beta - \alpha$ is contained in a candidate key for R . P177
- Let R be a relation schema, R_1 and R_2 from a decomposition of R . Decomposition is a 无损分解 if for all legal database instances r of R , $\Pi_{R_1}(r) \bowtie \Pi_{R_2}(r) = r$.
- In E-R model, an entity is represented by a set of 属性. A 关系 is an association among several entities.
- Assume relation r has b_r blocks and relation s has b_s blocks, therefore, in the best case, only $b_r + b_s$ block transfers would be required for $r \bowtie s$.
- An ideal hash function is uniform and 随机的, the former require that each bucket is assigned the same number of search-key values from the set of all possible values.
- To generate query-evaluation plans for an expression we have to generate logically equivalent expressions using equivalence rules.
- 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 $\lceil \log_{n+1} K \rceil$ 数据 事物可透明 并行结果等于串行结果
- A transaction has the following properties: atomic consistency, isolation and durability.
- When the final statement of a transaction has been executed, the transaction enters the partially committed state. After a transaction has been rolled back and the database has been restored to its previous state, the transaction enters the aborted state. 部分提交 活动 失败 中止

11. A schedule S is 可恢复 if a transaction T_j in S needs a data item previously written by a transaction T_i , then the commit operation of T_i appears before the commit operation of T_j . P375
12. 多值 attribute values or 复合 attribute values are not atomic. P193
13. A relation schema may have an attribute that corresponds to the primary key of another relation. The attribute is called a foreign key.
1. instance schema [phone_num] - 组值
2. 3NF nontrivial
3. lossless
4. Attribute relationship
5. Assume relation r has br block and relation s has bs blocks, therefore, in the best case, only $br+bs$ block transfers would be required for $r \bowtie s$.
6. 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.
7. To generate query-evaluation plans for an expression, we have to generate logically equivalent expression using equivalence rules.
8. 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 $\lceil \log_{[n/2]} K \rceil$.
9. A transaction has the following properties: atomicity, consistency, isolation and durability.
10. When the final statement of a transaction has been executed, the transaction enters the partially commite state. After a transaction has been rolled back and the database has been restored to its previous state, the transaction enter the aborted state.
11. A schedule S is recoverable if a transaction T_j in S reads a data item previously written by a transaction T_i , then the commit operation of T_i appears before the commit operation of T_j .
12. Mutivalued attribute values or composite attribute values are not atomic.
13. A relation schema may have an attribute that corresponds to the primary key of another relation. The attribute is called a foreign key.

adr < street 街址
city 城市

Part II [80 pts.] Answer the following question.

- [16 points] Database design I : Consider the following conditions

- i. The STUDENT may be taught by one and only one teacher . The TEACHER may be instructor of one or more STUDENT .
- ii. The TEACHER may be responsible for one and only one CLASS . The CLASS may be the responsibility of one and only one TEACHER .
- iii. The CLASS may be made of one of one or more STUDENT . The STUDENT must be a member of one and only one CLASS .
- iv. The CLASS must have one and only are 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 following attributes : TID ,TEACHERNAME , and TITLE .

- a) [8 points] Construct an E-R diagram showing these relationships .
- b) [4 points] Construct appropriate relation schemas for the above E-R diagrams .
- c) [4 points] Create an index *std_index* on the **student** relation with **SID** as the *search_key* .

表名和表头

2. [6 points] In database design , how to represent relationship set as relational schema ?

3. [14 points] Let $R = (A, B, C, D, E, F)$ be a relation with functional dependency $F = \{A \rightarrow CB, E \rightarrow FA\}$

- [2 points] Compute the candidate keys for R ;
- [6 points] Is R in 3NF ? If it is , justify your answer . If not , produce a decomposition of R into 3NF .
- [6 points] Is R in BCNF ? If it is , justify your answer . If not , produce a decomposition of R into BCNF .

A. Compute the candidate keys for R

Let us compute $E^+ E^+ = \{FACBE\}$

Let us compute $ED^+ E^+ = \{ABCDEF\}$

Thus $ED \rightarrow R$ ED is a superkey.

It is easy to see that $E \rightarrow /$ (不可推导) $R, D \rightarrow /R$

Thus, ED is a candidate key.

Note that ED is not implied by any other attributes. Thus any candidate key of R must contain ED . Further, ED is the unique candidate key of R

观察发现 E, D 不出现在左例

求 $\{ED\}^+ = \{ABCDEF\}$

故 ED 为 candidate key

B. Is R in 3NF? If it is, justify your answer. If not, produce a decomposition of R into 3NF.

Not. In $A \rightarrow CB$, A is not a superkey and is not contained in the candidate key.

Compute the canonical cover, we have $F_i = F$

According to F_i , we get $R_1 = (A, B, C)$, $R_2 = (A, E, F)$

Note none of schemas contains ED , we generate $R_3 = (D, E)$

Thus, decomposition of R

$R_1 = (A, B, C)$, $R_2 = (A, E, F)$, $R_3 = (D, E)$

解: 在 $A \rightarrow CB$ 中 ①

A 并非 superkey 且 A, C, B 不位于 candidate key 中 ②

故非 3NF

← 观察 $\{A \rightarrow CB, E \rightarrow FA\}$ 是正则范式

故 $R_1 = (A, B, C)$ $R_2 = (A, E, F)$

$R_3 = (D, E)$ (候选码)

C. Is R in BCNF? If it is, justify your answer. If not, produce a decomposition of R into BCNF.

Not. $A \rightarrow CB$ disobey the definition of BCNF.

$R_1 = (A, B, C)$, $R_2 = (A, D, E, F)$

In R_2 , $E \rightarrow FA$ disobey the definition of BCNF.

$R_3 = (A, E, F)$, $R_4 = (D, E)$

Thus, decomposition of R

$R_1 = (A, B, C)$, $R_2 = (A, E, F)$, $R_3 = (D, E)$

① $\alpha \rightarrow \beta$ 是平凡 ($\beta \subseteq \alpha$) (基本没用)

② $\alpha \rightarrow \beta$ α 是超键

在 $A \rightarrow CB$ 中 A 不是超键

分解: 把 $\alpha \rightarrow \beta$ 扒出, 删 β

$R_1 = (A, B, C)$, $R_2 = (A, E, F)$, $R_3 = (D, E)$

4. [28 points]

BOOK (Bookid , Title , Publishername)

BOOK_AUTHORS (Bookid , Authurname)

PUBLISHER (Publishername , Address , Phone)

BOOK_COPIES (Bookid , Branchid , No_Of_Copies)

LIBRARY_BRANCH (Branchid , Branchname , Address)

BOOK_LOANS (Bookid , Branchid , Cardno , DataOut , Duedata)

BORROWER (Cardno , Name , Address , Phone)

a) [16 points] 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" ?

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

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

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 " .

b) [3 points] Record the fact that the manager didn't maintain information about the book named " T&G " ,i.e. remove information about " T&G " .

A. Write appropriate SQL DDL statements for declaring the BOOK_AUTHORS relation.

Create table BOOK_AUTHORS

(
Bookid char(20),
Authurname char(200)
)

create Table book_authors!
book_id char(20),
author_name char(200),
primary key (book_id, author_name)
!

B. 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.

Temp $\leftarrow \Pi_{\text{Cardno}}(\text{BORROWER}) - \Pi_{\text{Cardno}}(\text{BOOK_LOANS})$

Res $\leftarrow \Pi_{\text{Name}}(\text{Temp} \bowtie \text{BORROWER})$ temp $\leftarrow \Pi_{\text{Cardno}}(\text{BORROWER}) - \Pi_{\text{Cardno}}(\text{BOOK_LOANS})$
res $\leftarrow \Pi_{\text{Name}}(\text{BORROWER} \bowtie \text{temp})$

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

res $\leftarrow \Pi_{\text{Title.Name, Address}}(\sigma_{\text{Branchname} = \text{"Sharpstown"} \wedge \text{DueDate} = \text{today}}(\text{BOOK_COPIES} \bowtie \text{LIBRARY_BRANCH} \bowtie \text{BOOK_LOANS} \bowtie \text{BORROWER}))$

II Title, Name, Address(σ Branchname="Sharpstown" \wedge DueDate=is today(LIBRARY_BRANCH \bowtie BOOK_LOANS \bowtie BORROWER \bowtie BOOK)

Select No_Of_Copies
from (BOOK natural join BOOK_COPIES) natural join LIBRARY_BRANCH
Where (Name = "Sharpstown" and Title=" ")

C. 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 No_Of_Copies
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

Select L.branchname, count(Bookid)
from BOOK_LOANS as B natural join LIBRARY_BRANCH as L
group by B.branch name

Q3: retrieve the names, address, and the 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

Select B.Name, B.Address, B.Cardno, count(*)
from ((BORROWER as B) natural join (BOOK_LOANS))
group by B.cardno
having count(1)>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, No_Of_Copies
From (((BOOK_AUTHORS natural join BOOK) natural join BOOK_COPIES) natural join LIBRARY_BRANCH)
Where Author_Name='Stephen King' and Branchname='Central'

D. Record the fact that the manager didn't maintain information about the book named "T&G", i.e. Remove information about "T&G".

Delete from BOOK_AUTHORS
Where Bookid in (select Bookid
From Book
Where Title='T&G')

Delete from BOOK_COPIES

Where Bookid in (select Bookid

From Book

Where Title='T&G')

Delete from BOOK_LOANS

Where Bookid in (select Bookid

From Book

Where Title='T&G')

Delete from BOOK

Where Title = 'T&G'

5. [16 points] Query Processing , Optimization and Transaction P317

- a) [4 points] please describe the implementation process of selection operation $\sigma_{A=c}(r)$, 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 . 最好情况: $n < d$
- b) [4 points] Describe the process of Indexed nested-loop join . 则查读 $h+1$ 个块
故 $(h+1)(t_c + t_r)$

辅助索引 (密)

对于外层关系的

每个元组t, 利用索引查找替代文件扫描

$b_r(t_r + t_s) + n_r \times c \leftarrow$ 单次嵌套代价

嵌套循环连接 P322

c) [4 points] Please describe the two-phase locking protocol and prove that it ensures conflict-serializable schedules and does not ensure freedom from deadlocks. P385

d) [4 points]

(1) 一个调度

	T_2	T_1
lock-x(A)		
		lock-x(B)
		lock-x(A)
lock-x(B)		

发生了死锁, 原因是 growing phase
可以同步进行

证: 当事务无关, 成立

有关且无死锁, 必定存在至少一个事务无依赖其它T
则其可以将读写提到其他T之前,
之后同理