北京邮电大学 2006——2007 学年第 2 学期

《数据库系统原理》期末考试试题（B）

考试注意事项	一、学生参加考试须带学生证或学院证明，未带者不准进入考场。学生必须按照监考教师指定座位就坐。 二、书本、参考资料、书包等物品一律放到考场指定位置。 三、学生不得另行携带、使用稿纸，要遵守《北京邮电大学考场规则》，有考场违纪或作弊行为者，按相应规定严肃处理。 四、学生必须将答题内容做在试题答卷上，做在试题及草稿纸上一律无效。 **五、填空题用英文答，中文答对得一半分。**

考试课程				考试时间		2007 年 6 月 25 日					
题号	一	二	三	四	五	六	七	八	九	十	总分
满分	9	11	6	10	20	15	6	6	8	9	
得分											
阅卷教师											

1. Fill in blanks. (1×9 points)

(1) Application programs are said to exhibit ___physical independency_____, if they do not depend on the physical schema, and thus need not be rewritten if the physical schema changes.

(2) DBS design can be divided into three stages, at the ___conceptual design_____ stage, the E-R model is used to describe the data objects in the world and the associations among the objects.

(3) An_____entity set_____ is a collection of entities of the same type.

(4) With respect to integrity constraints in DBS, ___referential integrity, *or* foreign key_____ constraints ensures that a value appearing in a relation for a given set of attributes also appears for a certain set of attributes in another relation.

(5) The commonly-used schemes of organization of records in files are heap file organization, ___ sequential file organization_____, and hashing file organization.

(6) With respect to ACID properties of the transaction, _isolation_____guarantees that, even though multiple transactions may execute concurrently, the system guarantee that all transactions seem to execute serially, so that the consistent states of DB can be preserved.

(7) The schedule S and S' are ___conflict equivalent_____if S can be transformed into *S'* by a series of swaps of non-conflicting instructions.

(8) With respect to handling of transaction deadlock, two commonly used schemes are deadlock prevention and deadlock____detection and recovery_____

(9).The several concurrent control strategies are lock-based protocols,____timestamp protocols, validation-based protocols and multiversion schemes.

答案：

（1）physical independency

(2) conceptual design

(3) entity set

(4) referential integrity, *or* foreign key

(5) sequential file organization

(6) isolation

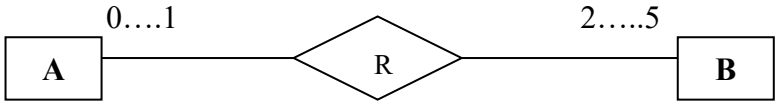
(7) conflict equivalent

(8) detection and recovery

(9) timestamp

2. Choice (1×11 points)

(1) For the E-R diagram given below, the mapping cardinality from **A** to **B** is___C_____



A. one-to-many B. one-to-one C. many-to-one D. many-to-many

(2) In SQL Select statement, the ***where*** subclause corresponds to the relational algebra operation D__

A. ∏ B. ρ C. ╳ D. σ

(3) Given the schema R(A, B, C, D, E, F, G) and the functional dependencies F={AE→D, CDFG→ AE, D → BF, CF→ AE} holding on it,___B_____ is a trivial functional dependency.(平凡函数依赖)

A. AE→D B. CDFG→ DG C. D → BF D. AE→ BF

(4) Given a relation r(R), which one of the following functional dependencies is **not** satisfied by r

___D_____

A. D→C B. BD→A C. C→D D. AB→D

A	B	C	D
4	4	1	1
6	5	4	9
3	4	2	3
6	5	5	2
9	1	0	5

(5) Given a schema R(A, B, C, D) and F={A→ B, B→C}, R is decomposed into R1(A, B, D) and R2(A, C, D), this decomposition is A, .

A. lossless B. lossy C. dependency preserving

(6) In SQL language, the statement that can be used for security control is C

A. commit B. rollback C. grant D. select

(7) With respect to the relation ***account***(*account_number*, *branch_name*, *balance*), all information except D belong to meta-data and are stored in the data dictionary.

A. the name and type of the attribute *branch_name*
B. the numbers of all the tuples in ***account***
C. the type of the index for ***account***, defined on *branch_name*
D. a tuple (A-101, downtown, 500) in ***account***

(8) Considering the measures of query cost, A is the predominant cost

A. disk accesses, i.e. accessing data from disk
B. CPU times to execute a query
C. communication costs, for example, in distributed or parallel DBS

(9) As for the following equivalence rules for transformation of relational expressions, which one is not right? B

A. $\sigma_{\theta_1 \wedge \theta_2}(E) = \sigma_{\theta_1}(\sigma_{\theta_2}(E))$

B. $\sigma_{\theta}(E_1 \times E_2) = E_1 \div E_2$

C. $E_1 \cup E_2 = E_2 \cup E_1$

□□□□□ D.□□□□□□□$\prod_L(E_1 \cup E_2) = (\prod_L(E_1)) \cup (\prod_L(E_2))$

(10) With respect to data access in DBS, the operation C is used to transfer the block in the disk buffer to the disk, and replaces the appropriate physical block here.

A. input(B_X) B. read(X) C. output(B_X) D. write(X)

(11) Consider the following concurrent schedules on transaction T1 and T2. Assuming TS (T1) < TS(T2), which one is **not** in accordance with timestamp protocol ? A .

A. S1 B. S2 C. S3

S1

T1	T2…
…	…
	read(X)
	write(X)
	read(Z)
write(Y)	
	write(Y)
read(Y)	
read(Z)	

S2

T1	T2
…	…
	read(X)
write(Y)	
	write(X)
	read(Z)
read(Y)	
	write(Y)
read(Z)	

S3

T1	T2
…	…
write(Y)	
	read(X)
read(Y)	
read(Z)	
	write(X)
	read(Z)
	write(Y)

答案：

(1) C

(2) D

(3) B

(4) D

(5) A

(6) C

(7) D

(8) A

(9) B

(10) C

(11) A

3. (6 points) Given relational schemas *Supplier*(S#, SNAME, CITY), *Part*(P#, PNAME, COLOR, WEIGHT), *Job*(J#, JNAME, CITY), *SPJ*(S#, P#, J#, QTY). In the relation *Supplier*, S# is the supplier identifier, SNAME is the supplier name, and CITY is the city where the supplier lives; In the relation *Part*, P# is the part code, PNAME is the part name, COLOR is the part color, and WEIGHT is the part weight; In the engineering relation *Job*, J# is the engineering identifier, JNAME is the engineering job name, CITY is the city where the engineering is located; A *part* can be supplied by some *suppliers*, and be used for some engineering *jobs*. The supplying relation *SPJ* describes the association among the objects *part*, *supplier* and *job*, QTY is the total number of the *parts* supplied.

Give relational algebra expressions for the following queries:

(1) Find the *names* of the suppliers who provide *red parts* for the engineering jobs located either in Beijing or Shanghai (3points)

(2) Find the *code*, *name* and *weight* of the part provided by the suppliers living in the city, at which the engineering job is also located and makes use of this part. (3points)

答案：

(1) $\Pi_{SNAME}(\sigma_{(Job.CITY='Beijing' \text{ or } Job.CITY='Shanghai') \text{ and } Part.COLOR='red'}(Supplier \bowtie Part \bowtie Job \bowtie SPJ))$ （3 points）

(2) $\Pi_{P\#, PNAME, WEIGHT}(\sigma_{supplier.CITY=Job.CITY}(Supplier \bowtie Part \bowtie Job \bowtie SPJ))$ (3 points)

4. (10 points) Considering the schema R=(athlete-id, game-event, grade, category of games, manger of games) that describes the sports meeting. It is assumed that

- for each athlete, if he takes part in a game event, he will achieve one and only one grade

- each game event belongs to one and only one category
- each category is managed by one and only one manager

这两个条件就是，虽然语句是被动句，但是前面的决定了后面的。

(1) According to the descriptions mentioned above, list the functional dependency set F that holds on R (3 points)

(2) List all the candidate keys of R. (2 points)

(3) Give a lossless and dependency-preserving decomposition of R into 3NF. (5 points)

答案：

函数依赖如下：

athlete-id, game-event→grade (1 point)

game-event→category of games (1 point)

category of games→manger of games (1 point)

候选键是：

athlete-id , game-event (2 points)

分解的 3NF 是：

(athlete-id, game-event, grade) (2 points)

（game-event ， category of games） (2 points)

（category of games, manger of games) (1 point)

5. (20 points) A pharmacies database needs to store the information about the pharmaceutical companies（制药公司）and others. The relevant information is as follows:

- Each pharmaceutical company is identified by its name and has a phone number.
- Each pharmacy (药房) is identified by its name and has an address, and several phone numbers.
- For each drug, the trade name and formula must be recorded, and the trade name identifies a drug uniquely.
- Each drug is developed by only one given pharmaceutical company, and one pharmaceutical company can develop several drugs and perhaps develops no dug now.
- Each pharmacy sells several drugs and has a price for each. A drug could be sold by several pharmacies, and the price may vary from one pharmacy to another.
- Pharmaceutical companies make contracts with pharmacies, but some pharmaceutical companies may have no contract. A pharmaceutical company can sign contracts with several pharmacies, and a pharmacy can also enter into contracts with several pharmaceutical companies. Each contract is described by a start date, an end date and the content of the contract.

(1) Design the E/R diagram for hospital database on basis of the information mentioned above. (10 points)

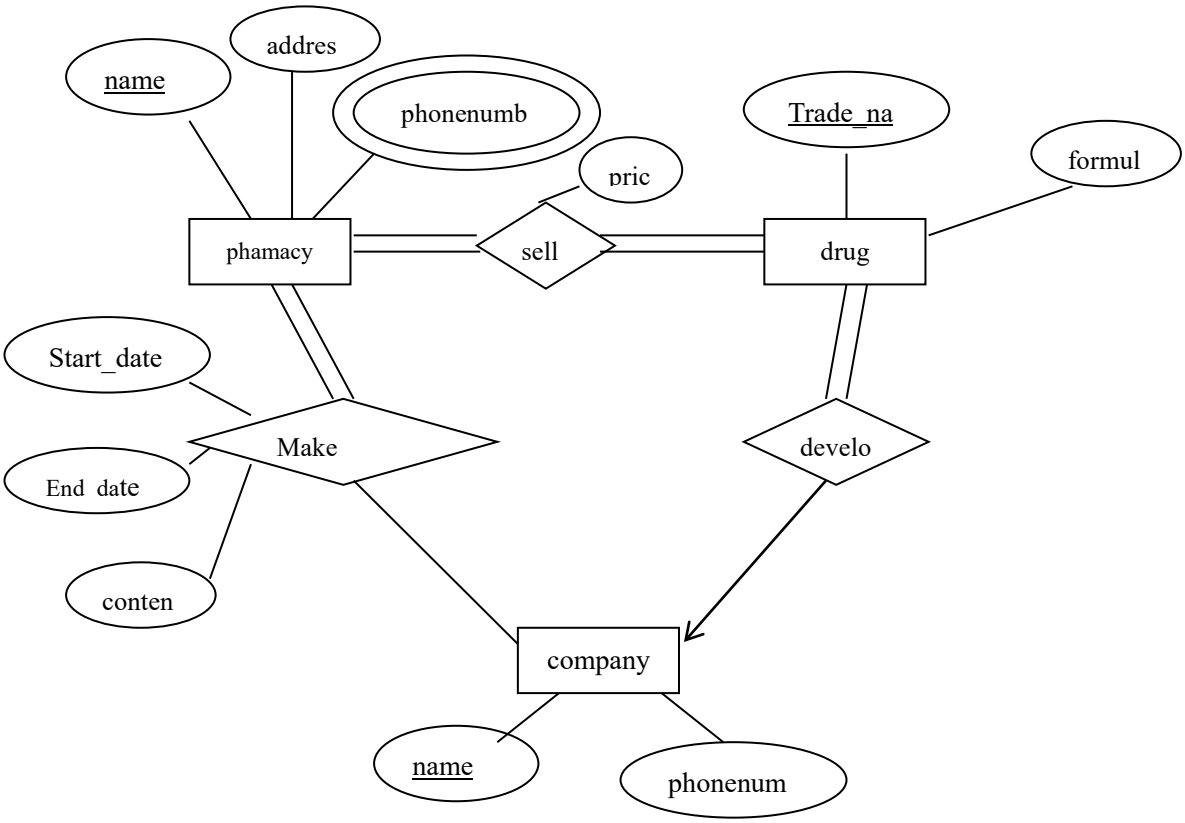
Note: the mapping cardinality of each relationship and participation of each entity to the

relationship should be described in the diagram.

(2) Convert the E-R diagram to the proper relational schemas, and give the primary key of each relation schema by underlines. (10 points)

答案：

（1）E-R 图



（2）转换后的关系模式

Pharmacy(name , address)

Pharmacy_phone(name , phonenumber)

Drug(Tade_name ,formular, company_name)

Company(name, phonenumber)

Sell(pharmacy_name,trade_name,price)

Contract(pharmacy_name, company_name, start_date,end_date,content)

其中 drug、phamacy 实体的参与联系答案不唯一。

评分标准：

ER 图有实体、属性、联系、主键、参与度、对应约束等得分点，错一个小点扣 0.5 分。

表分基本表、属性、主键等几个点，每一个小点错扣 0.5 分。

6. (15 points) There are three relations in the database, describing the employees and the companies that the employees work at,

Employee(employee-id, name, age, sex, e-city)

Company(company-id, company-name, company-city)

Works(employee-id, company-id, salary)

(1) The salaries of the employees, who live in Beijing and work at the company named SHOUGANG, increase by $100. Give a SQL statement to add this information into the database. (3 points)

(2) Give an SQL statement to find out the *name* and *id* of all the *employees*, who are male, work at the company named SHOUGANG, and whose *salaries* are below $2000. (3 points)

(3) Translate the SQL statement in (2) into an initial query tree, and give an optimized query tree for it, by means of heuristic query optimization. (9 points)

答案：

（1）
```
update    Works
set       salary = salary + 100
where     employee-id in    ( select    employee-id
                            from    Employee
                            where e-city='Beijing'
          and    company-id in    ( select    company-id
                                   from    Company
                                   where    company-name='SHOUGANG' )
```

或：
```
update    Works
set       salary = salary + 100
where     employee-id in    ( select    employee-id
                            from    employee as E, company as C, works as W
                            where    E.employee-id = W.employee-id      and
                            C.company-id = W. company-id      and
                            e-city ='Beijing' and company-name='SHOUGANG')
```

或：
```
update    Works
set       salary = salary + 100
where     exists    ( select    employee-id
                    from    employee as E, company as C, works as W
                    where    E.employee-id = W.employee-id      and
                             C.company-id = W. company-id      and
                             e-city ='Beijing' and company-name='SHOUGANG'
```

或：

```
with    employee-B-S (employee-id) as
            select    employee-id
            from    employee as E, company as C, works as W
            where    E.employee-id = W.employee-id      and
                        C.company-id = W. company-id      and
                          e-city =’Beijing’ and company-name=’SHOUGANG’
update    Works
set         salary = salary + 100
where Works.employee-id    in    employee-B-S
```

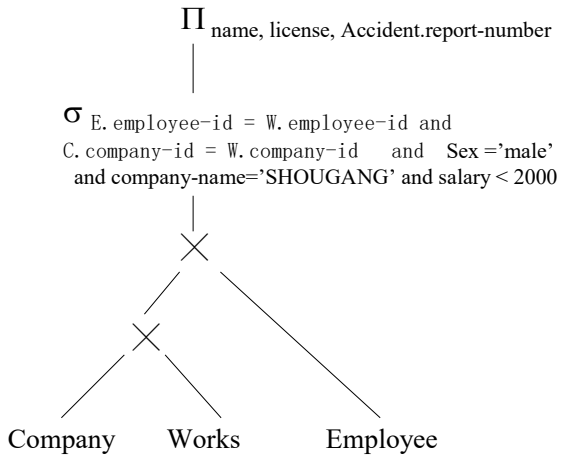
……要好好研究下 with 语句，是怎么回事，

评分标准：正确写出 Update 语句结构可得 1.5 分

(2) select name, employee.employee-id
from employee as E, company as C, works as W
where E.employee-id = W.employee-id and
C.company-id = W. company-id and
Sex =’male ’ and company-name=’SHOUGANG’ and salary < 2000

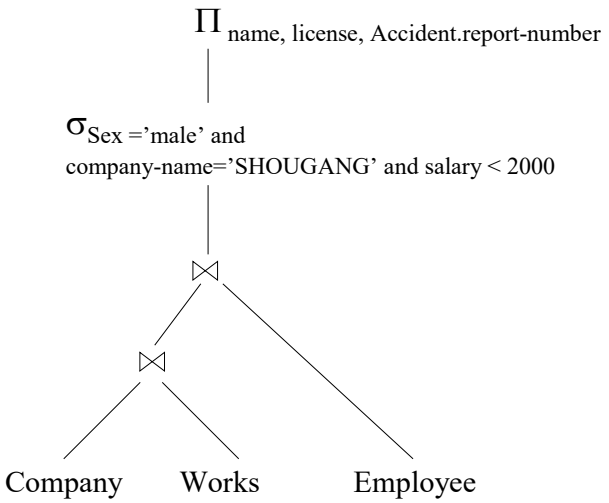
评分标准：漏掉连接条件扣 1 分

(3) 初始查询树



(a) initial query tree

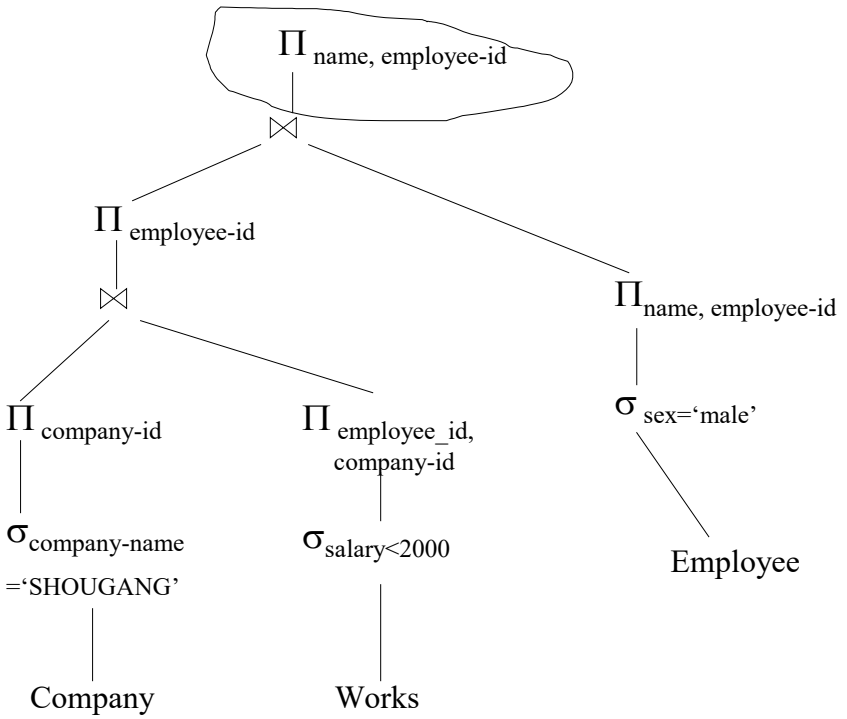
或：



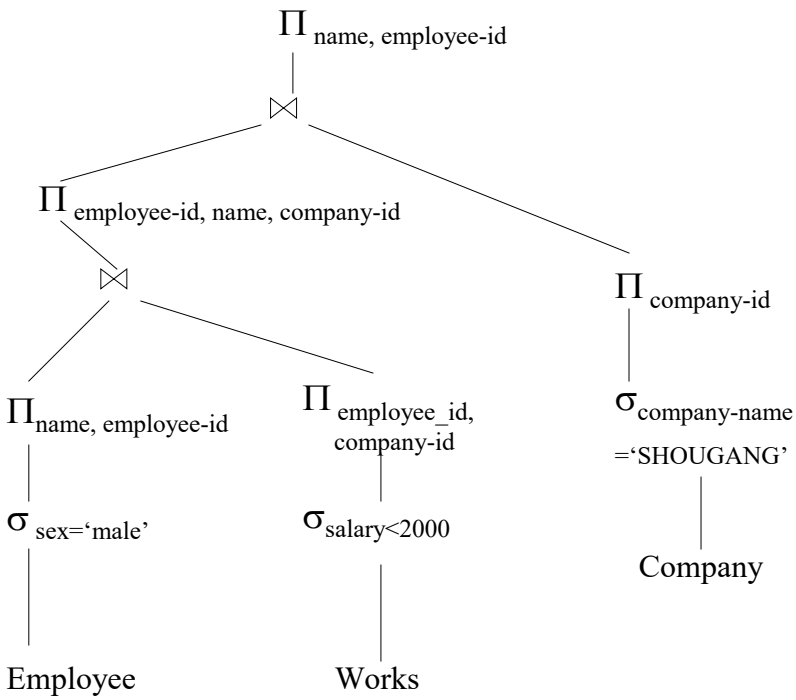
(a) initial query tree

评分标准：漏掉连接条件扣 1 分
叶结点关系表示错误扣一分

优化后的查询树：



或：

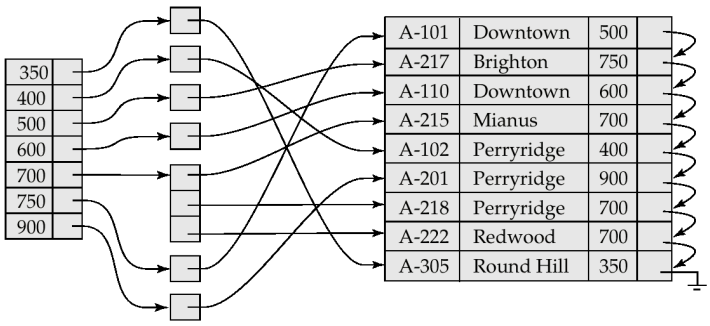


评分标准：每个选择、投影操作 0.5 分

7. (6 points) Given the data file ***account***(*account_number*, *branch_name*, *balance*) and the index defined on *balance*, as shown in the following figure,

(1) Is the index a dense or sparse index, why? (3 points)

(2) Is the index a clustering or non- clustering index, why? (3 points)



答案：

（1）稠密索引。因为对于被索引文件的 balance 属性上的每个值（共有 7 个），在索引中都有一个对应的索引项。

（2）non- clustering index。因为被索引文件中记录的排列顺序与索引项指定的查询顺序不一致。

评分标准：对各小题，正确回答是否得 1.5 分，正确阐述原因得 1.5 分。

8. (6 points) With respect to the concurrent schedule S on the transaction set {T1, T2, T3},

(1) Give the precedence graph G(S) for S (2 points)

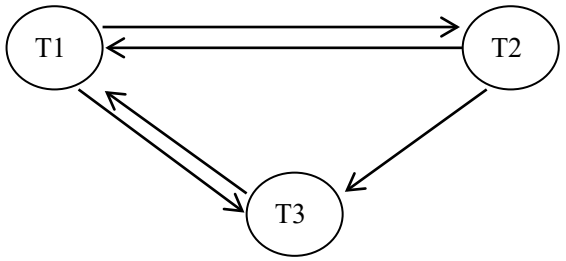
(2) Is S conflict serializable ? and why ? (4 points)

S

T1	T2	T3
read(P)		
	read(P)	
write(Q)		
	write(R)	
	write(Q)	
		write(Q)
		read(R)
write(R)		

答案：

（1）



评分标准：缺一条边扣 0.5 分

(2) 非冲突可串行的。因为前驱图中存在环路。

评分标准：正确回答是否冲突可串行得 2 分，正确阐述原因得 2 分。

9. (8 points) Considering the schedules in the following figures

(1) Is the concurrent schedule S1 shown in Fig.(a) under the two-phase locking protocol? and why ? (4 points)

(2) Is the concurrent schedule S2 shown in Fig.(b) a cascadeless schedule? and why ? (4 points)

S1

T_1	T_2
lock-S(A)	
read(A)	
lock-X(B)	
write(B)	
unlock(B)	
	lock-X(B)
	write(B)
unlock(A)	
	unlock(B)
	lock-X(A)
	write(A)
	unlock(A)

(a)

S2

T_1	T_2	T_3
write(A)		
commit		
	read(A)	
	write(B)	
		read(B)
	commit	
		write(B)
		commit

(b)

答案：

（1）不是符合两阶段锁协议的并发调度。因为事务 T2 在 unlock(B)后又 lock(A),不符合协议规定的锁增长和锁缩减的两个阶段。

（2）不是无级联回滚的调度。因为事务 T3 读了 T2 修改过但没提交的数据 B，T2 若在提交前失败回滚，将引起 T3 的级联回滚。

评分标准：

回答对第一问得两分。

10. (9 points) Considering the concurrent transactions T_1, T_2, T_3, T_4 and T_5, and the data items A, B, C and D modified by these transactions. It is assumed that

- the initial values of these data items are A=10, B=20, C=0, D=40.
- immediate database modification and checkpoint techniques are employed

With respect to the log in the following figure that describes the concurrent executing of T_1, T_2, T_3, T_4 and T_5, when a failure occurs, the log-based recovery scheme consults the log to determine the recovery operations (i.e. **redo**, **undo, ignore**) done on T_1, T_2, T_3, T_4 and T_5.

(1) Which transaction needs to be undone, which transaction needs to be redone? And which transaction can be ignored ? (5 points)

(2) After recovery operations on T_1, T_2, T_3, T_4 and T_5 are completed, what are the values of the data items A, B, C and D in the database? (4 points)

<T_1 start>

<T_1, A, 10, 20>

<T_2 start>

<T_3 start>

<T_2 , B, 20, 10>

<T_1 commit>

<T_3, C, 0, 10>

<T_3, C, 10, 30>

<checkpoint {T_2, T_3}>

<T_4 start>

<T_4, A, 20, 40>

<T_2, B, 10, 35>

<T_5 start>

<T_5, D, 40, 30>

<T_4, A, 40, 50>

<T_5, commit>

<T_4, commit>

Crash

答案：

（1） redo: T_4, T_5 (2 points)

undo: T_2, T_3 (2 points)

ignored: T_1 (1 point)

(2) A = 50 B=20 C=0 D=30 (4 points)