

主管
领导
审核
签字

哈尔滨工业大学 2023 学年春季学期

《数据库系统》试题

题号	一	二	三	四	五	六	七	八	九	十	总分
得分											
阅卷人											

片纸鉴心 诚信不败

一、单选题 (5 分, 每题 1 分)

- 若事务 T 对对象 A 已加 X 锁, 则其他事务对 A ()
A. 可以加 S 锁, 但不能加 X 锁 B. 不能加 S 锁, 但可以加 X 锁
C. 可以加 S 锁, 也可以加 X 锁 D. 不能加 S 锁, 也不能加 X 锁
- 下列聚集函数正确的是 ()。
A. SUM (*) B. COUNT (*) C. MAX (*) D. AVG (*)
- 设有部门和员工两个实体, 每个员工只能属于一个部门, 一个部门可以有多名员工, 则部门与员工实体之间的联系类型是 ()。
A. m:n B. 1:m C. 1:1 D. m:1
- 下列关于关系模式的叙述中, 错误的是 ()。
A. 关系中的一个属性可以对应于多个值域
B. 任何属性不可再分
C. 主键可以由多个属性组成
D. 关系中的属性具有顺序无关性
- 下面不属于事务的特性的是 ()。
A. 一致性 B. 隔离性 C. 完整性 D. 持久性

二、填空题 (8 分, 每空 1 分)

- 关系模型的三要素为_____、属性和域。
- 关系代数六种基本操作包括 _____, 投影, 并, _____, 笛卡尔积, 重命名。
- ER 图中三种主要元素为实体, 属性, _____。
- 2NF 是在 1NF 的基础上消除了非键属性对键属性的_____函数依赖, 3NF 是在 2NF 的基础上消除了非键属性对键属性的_____函数依赖。
- 设关系 R(A,B,C,D)和关系 S(C,D,E), 则自然连接 $R \bowtie S$ 用等价的关系代数基本运算可以表示为_____, 右外连接 $R \bowtie\!> S$ 用等价的关系代数基本运算可以表示为_____。

三、查询题（20 分）

已知一个关系数据库的模式如下：

- EMPLOYEE(E#, ENAME, SALARY, SUPER#, D#)记录员工数据，其中 E#是员工工号（字符串型），ENAME 是员工姓名（字符串型），SALARY 是员工薪资（浮点型），SUPER#是员工直接领导的工号（字符串型），D#是员工所属部门号（字符串型）；
- DEPARTMENT(D#, DNAME, MGR#)记录部门数据，其中 D#是部门编号(字符串型)，DNAME 是部门名称（字符串型），MGR#是部门领导的工号（字符串型）；
- PROJECT(P#, PNAME, D#)记录项目数据，其中 P#是项目编号（字符串型），PNAME 是项目名称（字符串型）D#是项目主管部门编号；
- WORKS_ON(E#, P#, HOURS)记录员工在项目工作的数据，E#是员工工号(字符串型)，P#是项目编号(字符串型)，HOURS 是一个员工在一个项目中的工作小时数（整型）。

注意：

（1）下划线标注的是对应关系模式的主码；（2）数据库中所有记录非空且不重复；（3）员工可以参与非本部门主管的项目中；（4）员工的工作小时数大于 0；（5）部门的员工平均工资=该部门总工资/该部门员工总数；（6）员工的小时平均工资=该员工工资/该员工工作总时长

请写出以下查询

- 1.（4 分）使用关系代数表达式查询“在 Research 部门工作且工资高于 8000 的员工姓名”。
- 2.（4 分）使用关系代数表达式查询“至少参与了项目编号为 P1 和 P2 的项目的员工工号”。
- 3.（4 分）使用 SQL 查询“每个部门的部门名称和员工平均工资”。
- 4.（4 分）使用 SQL 查询“没有参加项目编号为 P1 的员工姓名”。
- 5.（4 分）使用 SQL 查询“所有干过非自己部门主管任务的员工工号和员工姓名”。

姓名

学号

院系

密

封

线

四、计算与证明题（20 分）

设有关系模式 $R\{A,B,C,D,E\}$ ，其上有函数依赖集：

$F=\{AB\rightarrow CD, A\rightarrow DE, BD\rightarrow E, BE\rightarrow C, D\rightarrow A\}$

回答以下问题：

1. （5 分）求 BD 关于 F 的属性闭包。

2. （5 分）求所有候选键。

3. （5 分）求 F 的最小覆盖。

4. （5 分）把 R 分解为一组 3NF 关系模式，使得该分解既是无损连接分解，又保持函数依赖。

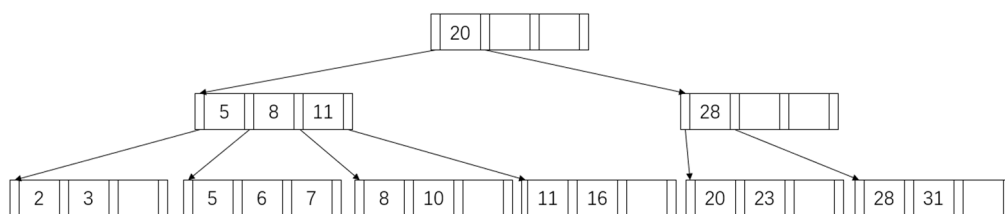
五、设计题（15 分，从题目 A 和题目 B 中任选一道作答）

题目 A：任选一个你熟悉的系统，正确分析该系统的需求，绘制相应的 ER 图。要求：(1) 简要叙述系统的需求。(2)系统中的实体数量不得少于 10 个。(3)必须包含一对一联系、一对多联系和多对多联系以及描述性属性。(4)将此 ER 图转换为关系模式。

题目 B：任选一个你熟悉的系统，正确分析该系统的需求，绘制出相应的 IDEF1X 图。要求：(1) 简要叙述系统的需求--注意 IDEF1X 图的完备性正确性评价以此需求为基本判断依据；(2) 实体数目不得少于 10 个；(3) 不得出现分类联系与分类实体；(4) 将此 IDEF1X 图转换为对应的关系模式。

六、算法题（6分）

1. (3分) 已知如下图所示的 B+树，绘制出该 B+树中删除索引项 23 后得到的 B+树。



2. (3分) 利用线性 hash 方法对以下记录进行 hash 存储：

5, 8, 19, 26, 30, 33, 36, 41

设 hash 函数 $h(x) = x \bmod 16$ ，请画出添加以上所有元素后，最终的索引结构以及关键步骤（进行桶的线性增长时）的索引结构。

➤ 注：线性 hash 表中最多容纳 $n \times b \times \theta$ 个记录， $\theta = 0.8$ ； n 为桶的数量，每个桶最多可容纳 b 个 hash 记录。

七、分析题（6分）

假设有两个关系表 $\text{Class}(\text{Class_ID}, \text{College}, \text{Teacher})$ 和 $\text{Student}(\text{Student_ID}, \text{Name}, \text{Class_ID})$ ，其中 Class 有 10000 个元组，每块可以容纳 100 个 Class 元组； Student 有 5000 个元组，每块可以容纳 20 个 Student 元组。使用基于块的嵌套循环连接算法，将 Class 和 Student 连接起来，缓存区可用的内存页数 $M=11$ ，并分析以下问题：

1. (2 分) 内外关系分别是哪个表?
2. (2 分) 该连接算法的 I/O 代价是多少?
3. (2 分) 是否可以使用排序归并连接算法连接 Class 和 Student, 并说明理由?

八、分析题（8分）

设员工项目管理数据库有如下三个关系模式：

- EMPLOYEE(E#, ENAME, SALARY)记录员工数据，其中 E#是员工工号，ENAME 是员工姓名，SALARY 是员工薪资；
- PROJECT(P#, PNAME)记录项目数据，其中 P#是项目编号，PNAME 是项目名称；
- WORKS_ON(E#, P#, HOURS)记录员工在项目工作的数据，E#是员工工号，P#是项目编号，HOURS 是一个员工在一个项目中的工作小时数。

其中，E#、P#分别是 EMPLOYEE、PROJECT 表的主码，(E#, P#)是 WORKS_ON 表的主码，也分别是参照 EMPLOYEE、PROJECT 表的外码

用户有一查询语句：

```
Select ENAME
```

```
From EMPLOYEE, PROJECT, WORKS_ON
```

```
Where WORKS_ON.E#=EMPLOYEE.E# and
```

```
WORKS_ON.P#=PROJECT.P# and
```

```
PNAME="DBMS"
```

检索参与“DBMS”项目的员工的姓名

1.（2分）写出以上 SQL 语句所对应的关系代数表达式。

2.（6分）画出上述关系代数表达式所对应的查询计划树。使用启发式查询优化算法，对以上查询计划树进行优化，并画出优化后的查询计划树。设 WORKS_ON 表有 10000 条元组，PROJECT 表有 20 条元组，EMPLOYEE 表中有 1000 条元组，WORKS_ON 中满足参与“数据库系统”项目的元组数为 50，计算优化前与优化后的查询计划中每一步所产生的中间结果大小。

九、分析题（6分）

1. （2分）考虑如下三个事务的调度，这是否是一个冲突可串行化调度？请解释一下你的判断依据。如果是，将它转换成一个串行调度。

T_0	T_1	T_2
$r_0(A)$		
$w_0(A)$		
		$r_2(A)$
		$w_2(A)$
	$r_1(A)$	
$r_0(B)$		
		$r_2(B)$
$w_0(B)$		
	$r_1(B)$	$w_2(B)$

2. （4分）考虑如下两个事务：

T_1 :

read(A, t);

t:=t+1;

write(A, t);

read(B, t);

t:=t+1;

write(B, t).

T_2 :

read(A, s);

s:=s*2;

write(A, s);

read(B, s);

s:=s*2;

write(B, s).

(1) （2分）在 T_1 和 T_2 上分别添加加锁和解锁指令，使其遵从两段锁协议。

(2) （2分）基于(1)给出的两段锁协议，是否存在一个发生死锁的并发调度序列，为什么。

十、分析题（6分）

设一个数据库系统启动后中，执行4个事务T0、T1、T2和T3。四个事务的内容如下：

T0: $A := A / 5$ (读入数据库元素A的值，除以5后，再写回A的值)

T1: $B := B + 70$ (读入数据库元素B的值，加上70后，再写回B的值)

T2: $C := C - 80$ (读入数据库元素C的值，减去80后，再写回C的值)

T3: $D := D * 2$ (读入数据库元素D的值，乘以2后，再写回D的值)

除了这四个事务外，系统中无其他事务执行。设四个事务开始前，数据库元素A、B、C、D的值分别为 $A=100$ ， $B=-20$ ， $C=60$ ， $D=40$ 。在执行这四个事务的过程中，系统发生了故障。系统重启后，经故障恢复，数据库元素A、B、C、D的值被恢复为 $A=20$ ， $B=50$ ， $C=-20$ ， $D=80$ 。故障恢复时，数据库系统日志文件中包含如下日志记录，这里只给出部分日志记录。已知该数据库管理系统使用基于undo-redo日志的故障恢复技术，这段日志中有且仅有1个检查点。在数据库恢复中 $Undo=\{T3\}$ ， $Redo=\{\emptyset\}$ 。

请根据以上信息将下列日志补充完整（任选其一补充即可）：

01: <T0, START>	01: <T0, BEGIN>
02: _____	02: _____
03: <T1, START>	03: <T1, BEGIN>
04: <T0, COMMIT>	04: <T0, COMMIT>
05: _____	05: _____
06: <T1, B, -20, 50>	06: <T1, B, -20, 50>
07: _____	07: _____
08: <T2, COMMIT>	08: <T2, COMMIT>
09: <T1, COMMIT>	09: <T1, COMMIT>
10: <T3, START>	10: <T3, BEGIN>
11: _____	11: _____
12: <T3, D, 40, 80>	12: <T3, D, 40, 80>
	13: <END CHECKPOINT>