

数据库系统

试题卷 (A)

| 题号 | 一 | 二 | 三 | 四 | 五 | 六 | 七 | 八 | 九 | 总分 |
|----|---|---|---|---|---|---|---|---|---|----|
| 分数 | | | | | | | | | | |

一、填空题（每空 1 分，共 10 分）

- 1、实体联系模型，也称为 E-R 模型，的三个主要元素为 (a)、(b)、(c)。
- 2、基于锁的并发控制协议中，常用锁的类型有 (d) 和 (e)。
- 3、RAID1、RAID4 和 RAID5 中每块磁盘都承担冗余数据存储的是 (f)。
- 4、数据库的三级模式架构中包含 2 级映射，其中 2 级映射包括 (g) 和 (h)。
- 5、建立在函数依赖基础上的关系模式规范化形式主要有：第一范式（1NF）、第二范式（2NF）、第三范式（3NF）、(i)。
- 6、事务的 ACID 特性是指原子性、一致性、隔离性和 (k)。

二、选择题（每题 1 分，共 5 分）

- 1、关于关系模型，下列说法不正确的是_____。
 - (A) 属性的次序是可以交换的
 - (B) 不能有重复的元组
 - (C) 属性是可再分的
 - (D) 不同属性的值可来自相同的域
- 2、数据库系统日志文件用于保存_____。
 - (A) 程序执行的过程
 - (B) 所有数据操作
 - (C) 程序执行的结果
 - (D) 数据库更新操作
- 3、关于数据模型的说法，正确的是_____。
 - (A) 数据库设计过程中只需使用一种数据模型
 - (B) 实体-联系（ER）模型用于将现实世界抽象为关系
 - (C) 在将 ER 模型转换为关系数据库模式时，实体型会被转换为关系模式

(D) 由一个 ER 模型转换成的关系数据库模式是唯一的

4、关于可扩展外存哈希表，以下说法不正确的是_____。

- (A) 当需要增加桶的数量时，可扩展哈希表每次将桶的数量加倍
- (B) 不同桶中的元组可以存储于同一个磁盘块
- (C) 可扩展哈希表支持 ORDER BY 查询
- (D) 可扩展哈希表不支持范围查询

5、有两个关系 R 和 S，分别包括 15 个和 10 个元组，则在 $R \cup S$ ， $R - S$ ， $R \cap S$ 中可能出现的元组数目是_____。

- (A). 15, 10, 10
- (B). 18, 7, 7
- (C). 21, 11, 4
- (D). 25, 13, 2

三、问答题（每题 2 分，共 10 分）

- 1、数据库的模式和实例有什么区别？
- 2、和基于文件系统的数据管理方法相比，数据库系统有哪些主要优点？
- 3、在数据库系统中，数据独立性有什么作用？
- 4、在设计关系数据库时，关系数据库模式的规范化程度越高越好吗？
- 5、SQL 语言中 INNER JOIN 和 OUTER JOIN 有什么区别，列举三种类型的 OUTER JOIN。

四、解答题（共 20 分）

在图书管理数据库中，有如下三个关系表：

B(B#, BNAME, AUTHOR, TYPE)

S(S#, SNAME, CLASS)

L(S#, B#, DATE)

其中：

B 为图书信息表：B#为图书编号，BNAME 为书名，AUTHOR 为作者，TYPE 为图书类别；

S 为学生信息表：S#为学号，SNAME 为学生姓名，CLASS 为班级号；

L 为借阅信息表：S#为借阅人学号，B#为被借阅图书编号，DATE 为借阅日期。

备注：限定同一本书(即同一编号的书)每人永远只能借一次。另外，当有多本同名书籍时图书编号各不相同。

请使用 SQL 语言回答下列问题:

- (1) 检索每种类型图书的总数, 输出图书类型及相应类型的图书总数。(4 分)
- (2) 查询借了"Eclipse 入门"但没有借"JAVA 编程思想"的读者, 输出其学号。(4 分)
- (3) 找出借书超过 5 次的读者, 输出其学号、姓名及所借图书次数(注: 同一学号的学生若借阅不同编号的同名书籍, 则算作多次借阅)(4 分)
- (4) 建立一个视图, 显示班号为"1103101"班的学生们的借书信息(只要求显示姓名和书名)(4 分)
- (5) 找出被班号为"1103101"的所有学生借阅过的图书类型。(4 分)

五、解答题 (15 分)

二、考虑关系模式 $R(F, G, H, I, J)$, 其函数依赖集为 $FD = \{F \rightarrow I, J \rightarrow I, I \rightarrow G, FH \rightarrow I, GH \rightarrow I, IH \rightarrow F\}$ 。回答下列问题。

- (1) 使用 Armstrong 公理证明 $JH \rightarrow F$ 被 R 的函数依赖集 FD 逻辑蕴含。(3 分)
- (2) 给出 R 的所有候选键, 并用属性集闭包证明其为候选键。(3 分)
- (3) 关系模式 R 属于哪个范式? 给出理由。(3 分)
- (4) 证明 R 的函数依赖集 FD 与函数依赖集 $FD' = \{F \rightarrow I, J \rightarrow I, I \rightarrow G, GH \rightarrow I, IH \rightarrow F\}$ 等价。(3 分)
- (5) 将关系模式 R 分解为一组 3NF 关系模式, 保证该分解满足无损连接性和函数依赖保持性。(3 分)

六、分析题 (15 分)

数据库中包含 4 个关系 $R(a, b)$, $S(b, c)$, $T(c, d)$, $U(d, a)$ 。这 4 个关系的统计信息如下:

- R 中有 1000 条元组, R 中不同的 a 属性值个数为 100, 不同的 b 属性值个数为 200;
- S 中有 1000 条元组, S 中不同的 b 属性值个数为 100, 不同的 c 属性值个数为 500;
- T 中有 1000 条元组, T 中不同的 c 属性值个数为 20, 不同的 d 属性值个数为 50;
- U 中有 1000 条元组, U 中不同的 d 属性值个数为 1000, 不同的 a 属性值个数为 50;

请回答下列问题。

- (1) 考虑连接顺序 $((R \bowtie S) \bowtie T) \bowtie U$, 估计每个自然连接操作结果的元组数量(共计三次自然连接)。最后说明该查询的估计代价是多少。(5 分)

(2) 关系 R 和 S 均采用聚集存储, 每个磁盘块最多可存储 50 条 R 的元组或 25 条 S 的元组。设缓冲区中有 11 个磁盘块可用。使用嵌套循环连接(nested loop join)算法执行 $R \bowtie S$ 需要执行多少次磁盘 I/O? (不计算保存连接结果所需的磁盘 I/O 次数, 也不考虑缓存连接结果所需要的缓冲区页面) (5 分)

(3) 关系 T 和 U 均采用聚集存储, 每个磁盘块最多可存储 20 条 T 的元组或 10 条 U 的元组。设缓冲区中有 20 个磁盘块可用。使用排序归并连接(sort merge join)算法执行 $T \bowtie U$ 需要执行多少次磁盘 I/O? (不计算保存连接结果所需的磁盘 I/O 次数, 也不考虑缓存连接结果所需要的缓冲区页面) (5 分)

七、解答题 (10 分)

(1) 设 $r_i(X)$ 与 $w_i(X)$ 分别表示事务 T_i 读、写数据单元 X , 则一个并发调度可以抽象为读、写串。基于上述表示, 请判断下面两个并发调度是否是冲突可串行化的, 为什么? (5 分)

调度 S: $r_2(A); r_1(B); w_2(A); r_3(A); w_1(B); w_3(A); r_2(B); w_2(B);$

调度 S': $r_2(A); r_1(B); w_2(A); r_2(B); r_1(A); w_1(B); w_1(A); w_2(B);$

(2) 已知系统使用基于日志恢复数据库的即时更新协议, 如果系统在写入日志记录 $\langle T_0, B, 2050, 2100 \rangle$ 后即发生故障, 请给出上述事务哪些需要 REDO、哪些需要 UNDO, 并给出数据库恢复后数据项 A、B、C 的值。(5 分)

| |
|---|
| 1: $\langle T_0, \text{start} \rangle$ |
| 2: $\langle T_0, B, 2000, 2050 \rangle$ |
| 3: $\langle T_1, \text{start} \rangle$ |
| 4: $\langle T_1, C, 700, 600 \rangle$ |
| 5: $\langle T_1, \text{commit} \rangle$ |
| 6: $\langle T_2, \text{start} \rangle$ |
| 7: $\langle T_2, A, 500, 400 \rangle$ |
| 8: $\langle T_0, B, 2050, 2100 \rangle$ |
| 9: $\langle T_0, \text{commit} \rangle$ |

八、解答题 (5 分)

我们使用基于可扩展散列(extensible hashing)的索引来支持对数据的快速访问。给定查找键值 k , 设散列函数值 $h(k)$ 是一个 b 位二进制数, 我们用 $h(k)$ 的最后 d 位作为查找键值为 k 的记录在散列表中所在的桶的桶号。设该散列表中数据记录的查找键为整型数, $h(k) = k \bmod 12$, 散列表中每个桶最多容纳 2 个记录。该散列表最初有 2 个桶, 编号分别为 0 和 1, d 的初始值为 1。我们准备按如下查找键顺序向散列表中添加数据记录:

4, 15, 18, 26, 36, 42

当全部数据记录都插入完毕后, 该索引处于什么状态? 画出当时的索引, 数据记录用其查找键值表示即可, 用二进制数标明桶号, 注明 d 的值。

九、设计题（10 分）

工厂需建立一个管理数据库存储以下信息：

- 1) 工厂：厂名、厂长姓名；
- 2) 车间：车间号、车间主任姓名、地址、电话；
- 3) 仓库：仓库号、仓库主任姓名、电话；
- 4) 零件：零件号、重量、价格；
- 5) 产品：产品号、价格；

上述实体存在如下联系：

- a) 一个工厂内有多个车间和多个仓库，一个车间或一个仓库都只能属于一个工厂；
- b) 一个车间生产多种产品，每种产品只能产自一个车间；
- c) 一个车间生产多种零件，一种零件也可能为多个车间所制造；
- d) 一个产品由多种零件组成，一种零件也可装配出多种产品；
- e) 产品和零件均存入仓库。

根据上述要求，完成如下工作：

（2）画出该系统的 E-R 图（5 分）；

（2）将该 E-R 图转换为相应的关系模式，并标出主码（5 分）。