

# 复旦大学计算机科学技术学院

## 2022-2023 第一学期《数据库引论》期末考试试卷

### A 卷 共 5 页

课程代码: COMP130010.01

考试形式: ☐开卷 ☒闭卷

2022 年 12 月

(本试卷答卷时间为 120 分钟, 答案必须写在试卷上, 做在草稿纸上无效)

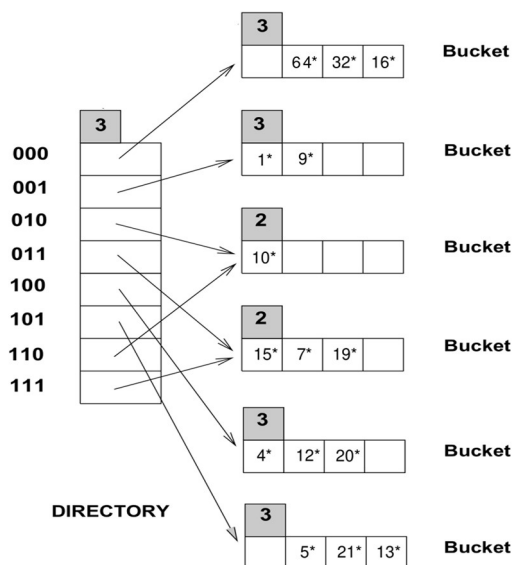
专业\_\_\_\_\_学号\_\_\_\_\_姓名\_\_\_\_\_成绩\_\_\_\_\_

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

#### 一、单项选择题 (每题 2 分, 共 12 分)

- DB 的三层模式结构中, 最接近外部存储器的是: ( )  
A、子模式                  B、外模式                  C、概念模式                  D、内模式
- 下列关于关系模式 R 的说法中, 正确的是: ( )  
A、若 R 属于 3NF, 则它一定消除了插入和删除异常。  
B、若 R 属于 2NF, 则 R 中一定不存在非主属性对关键码的局部依赖。  
C、如果 R 是 BCNF, 则其一定是无损分解。  
D、若 R 的所有非主属性对每一个码都是完全函数依赖, 则 R 属于 4NF。
- 下面关于数据库设计的相关说法中, 错误的是: ( )  
A、参与一个联系中的实体数目称为映射基数。  
B、数据库设计分为概念设计、逻辑设计、物理设计、实施与调试四个阶段。  
C、在概念设计阶段, 需要建立数据字典。  
D、从子类到超类的抽象化过程称为概化。
- 以下关于锁的说法中错误的是: ( )  
A、“两阶段封锁”指的是释放锁之后再获得锁要分成两个阶段, 不能同时进行。  
B、在事务依赖图中, 如果存在环那么则可能产生“死锁”。  
C、封锁带来的“死锁”问题可以用“定期检测和解除”方式解决。  
D、封锁对象的大小称为封锁的“粒度”, 封锁对象可以是属性值还可以是物理页。
- 下图是一个可扩充哈希表插入若干个值之后的状态, 假设这个哈希表从最初时间点开始到现在都没有进行过删除操作, 那么导致 Bucket 最后一次分裂的插入值可能是哪个: ( )

(装订线内不要答题)



- A、10                      B、19                      C、9                      D、64

6. 下图是一个事务发生故障（crash）的执行历史，用 T 表示事务，用 P 表示内存中的页，如果数据库采用的是 no-force 策略，则下列说法错误的是：（ ）

LSN	LOG
10	update: T1 writes P5
20	update: T2 writes P3
30	T2 commit
40	T2 end
50	update: T3 writes P1
60	update: T3 writes P3
	✗ CRASH, RESTART

- A、在发生故障时，P3 有可能是一个脏页。  
 B、T2 提交时，T2 修改的 P3 已经被写回磁盘。  
 C、在事务故障恢复过程中，T2 需要进行 Redo 操作。  
 D、在事务故障恢复过程中，T3 需要进行 Undo 操作。

二、不定项选择题（下列每题的各选项中，有一个或多个选项正确。每题 4 分，共 12 分。少选给 2 分，多选或错选不给分。）

1. 以下关于物理存储介质的说法中正确的是：（ ）  
 A、第三级存储器包含光盘、磁带、胶片等设备。  
 B、机械硬盘读取数据时，磁头转动，盘片不动。  
 C、磁盘的 checksum 用于检测磁盘是否正确进行读写，读写时各算一遍。

D、磁盘冗余磁盘阵列（RAID）的所有级别中，Level 0 是最适用于高可靠性系统。

2. 以下关于存储结构的说法中正确的是：（）

A、ISAM 索引是静态索引，只有叶节点中溢出的节点才会变化，因此相比 B+树索引并发性更好。

B、相比 B+树索引，哈希索引常用于范围查询。

C、一般的缓冲区管理器是一个专用硬件，用于将数据库所需要的页从硬盘中载入内存。

D、对于线性哈希，它是一种动态哈希方式，由一个 bucket 的序列构成，没有 directory。

3. 以下关于事务的说法中，错误的是：（）

A、事务的 ACID 性质包括原子性、一致性、隔离性和持久性。

B、若事务 T 对数据对象 A 加上 X 锁，则事务 T 可以读 A，其他事务能对 A 加 S 锁。

C、在读已提交的隔离级别下，事务 B 在事务 A 修改过数据之后，就可以读取到 A 修改的数据。

D、数据库中，在检查点可能会进行写入硬盘操作。

### 三、解答题（76 分。其中第 1 题 16 分，第 2 题 10 分，第 3 题 5 分，第 4 题 13 分，第 5 题 14 分，第 6 题 8 分，第 7 题 10 分）

1. 假设员工数据库中有 3 个关系：

职工关系 EMP (E#, ENAME, AGE, SEX, ECITY)

工作/兼职关系 WORKS (E#, C#, SALARY)

公司关系 COMP (C#, CNAME, CITY)

使用 **SQL 语句** 写出以下操作（每题只能用一条语句作答，并给出必要文字说明）：

(1) 使用 **元组表达式** 写出：检索工作在公司名为“A”公司的员工的姓名和年龄（2 分）。

(2) 使用关系代数表达式写出：查询 ABC 公司中所有薪水高于 10000 的男性员工的员工姓名与薪水。并画出语法树（4 分）

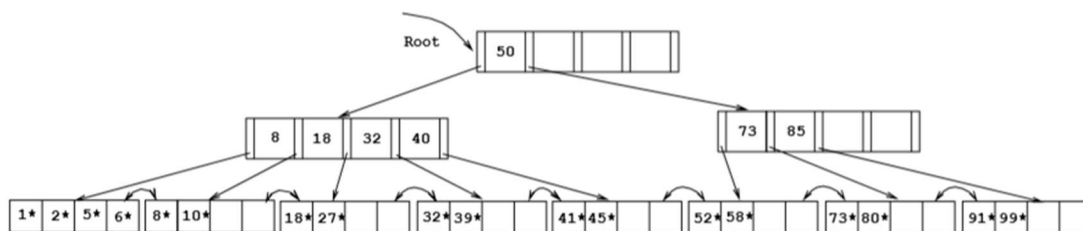
(3) 请写出工作关系 WORKS 的建表语句，注意指出主键和外键。（2 分）

(4) 查询所有工作城市和公司城市不同的员工姓名与工作城市。(2 分)

(5) 查询所有工资小于平均工资的女员工的员工编号、姓名与薪水(3 分)

(6) 查询这样的员工工号和姓名：该员工至少在员工 E6（员工号）兼职的所有公司工作(3 分)

2. 考虑如下秩为 2 的 B+树，请按照要求回答下面的问题（完整画出最终结果，并保留必要中间步骤，如节点分裂、合并前后）：



(1) 将码值为 3、9 的数据项分别插入原始树，画出插入之后的 B+树。(3 分)

(2) 将码值为 8、1 的数据项按顺序从原始树中删除，画出操作后的树（对于可能的重分布，先检查其左兄弟节点）(3 分)

(3) 在（1）插入值 3 的过程中，分别需要多少页的读操作与多少页的写操作？请给出理由（假设目前内存有足够缓冲区）(4 分)

3. 考虑一个关系，该关系包含的记录数为 1,000,000。已知每个页（page）大小为 1024B，rid 长为 8B，每条记录的主键长为 32B，每条记录总长 128B。主键上建有 B+树索引，假设 B+树（除根节点外节点）的占满率为 60%，要求使用方法 II（即每个数据项形如：<key, rid>）建立簇聚 B+树索引（**存储记录的页尽量存满**），则共需要多少页存储（B+树索引和记录）？（**要求给出具体的操作过程和必要的说明，仅给出答案不得分**）（5 分）
4. 考虑员工和公司两个关系：（员工信息） $R$  和（公司信息） $S$ ，其中  $R$  共有 10,000 条记录，每页包含 10 个记录； $S$  共有 5000 条记录，每页包含 10 个记录。对于连接操作  $R \bowtie_{R.a=S.b} S$ （可根据需求任选  $R$  或  $S$  为外关系），其中属性  $b$  为  $S$  的主键。假设有 36 个可用的缓冲区页， $R$  和  $S$  上没有建立任何索引，且  $R$  和  $S$  均存储于简单堆文件。在不考虑输出查询结果所需 I/O 操作的情况下，请回答以下问题。（**要求给出具体的操作过程和必要的说明，仅给出答案不得分**）
- (1) 使用“嵌套循环连接算法（tuple-at-a-time）”，“块嵌套循环连接算法”进行  $R$  和  $S$  的连接操作，请分别计算它们至少需要多少次 I/O 操作，以及如果要保持开销不变，至少需要多少个缓冲区页。（4 分）
- (2) 当采用“哈希连接”（分为 Partition 阶段和 Probing 阶段）的方法进行  $R$  和  $S$  表的连接操作，请计算需要多少次 I/O 操作？要确保这样的哈希连接能成功进行至少需要多少个缓冲区页？（若为  $N$  页数据建立内存哈希表，需要  $f \times N$  个缓冲区页，其中  $f$  称为“经验系数”或称为“数据分布的偏差参数”）（3 分）
- (3) 如果使用(2)中的哈希索引进行  $S$  和  $R$  的左外连接（即  $S \text{ NATURAL LEFT OUTER JOIN } R$ ）请描述出整个过程。（2 分）

(4) 对于员工信息表，我们有如下表结构，

R (eid, ename, title, phone, address)

对于 R 的所有属性都是字符串类型，且长度都是一致的，eid 是候选键，对于查询：

Select title, count(\*) from R group by title;

假设现在我们只有 25 个可用的缓冲区页，其它条件不变，请使用一般的归并排序算法完成上面的查询操作，并计算出所需要的 IO 次数。（4 分）

5. 下面是一个课程教师信息关系 R

R(tid: char(100), tname: char(50), tbirthday: char(50), title:char(100), department:char (100))

在关系 R 中，存在以下索引（均使用方法 II，即每个数据项形如：<key, rid>，rid 大小忽略不计）：tid 上的哈希索引，tname 上的非聚簇 B+树索引，tbirthday 上的哈希索引，(title,department)上的聚簇 B+树索引，title 上的哈希索引。为简化问题，假定上面的索引同时建立，哈希表与 B+树的占满率均为 100%，控制信息、指针大小均忽略不计，数据文件尽量存满。

已知关系 R 中，tid 为候选键，每条纪录的大小为 400 字节，关系 R 的数据文件总共有 10,000 页，数据文件中每页包含 10 条记录。假设在 B+树中单次查找适当的叶节点的 I/O 开销为 2，在哈希表中单次查找适当的桶的开销为 1.2。在不考虑输出查询结果所需 I/O 操作的情况下，考虑以下查询，选用何种读取路径最好？至少需要多少次 I/O？（要求给出具体的操作过程和必要的说明，仅给出答案不得分）

(1) select tid, tname from R where title = 'SeniorProfessor'，假设 title 此满足条件的元组数量占总数量的 10%，请计算出所有可能的答案，并取最小代价的方案（4 分）

(2) select tid, tname from R where title = 'Manager' or tname like 'Taylor%'，假设 title 满足条件的有 5%，tname 满足条件的有 10%，不存在同时满足条件的记录。请计算出所有可能的答案，并取最小代价的方案。（6 分）

(3) select title,count(\*) from R where tbirthday > "2002-04-05" and department like 'Person%' and title like 'Senior%' group by title（假设 title 满足条件的有 20%，department 满足条件的有 20%，tbirthday 满足条件的有 10%，title 和 department 同时满足的有 10%。三者同时满足条件的有 5%）（4 分）

6. 简要回答以下问题：

(1) 在概念设计中，如何把多值属性变换成系统容易实现的形式？（4分）

(2) 多个事务并发执行时，如果系统不加控制会破坏数据库的数据一致性，请举例出数据库的并发操作会带来的2个问题，并给出解决办法。（4分）

7. 考虑对以下事务 T1, T2 和 T3 的并发调度。

数据库中 A 的初始值为 100；读操作（FIND）表示从数据库中读值，括号内显示读取到的值；写操作（:=）修改并更新对象的值。

时间	事务 T1	事务 T2	事务 T3	事务 T4
t1	FIND A (100)	FIND A (100)		
t2	A := A - 30		FIND A (100)	
t3	*COMMIT*			
t4		FIND A (70)		FIND A (70)
t5		A := A + 50	A := A * 2	
t6		*COMMIT*		
t7			*ROLLBACK*	
t8				A := A * 3
t9				
t10				*COMMIT*

(1) t10 以后，A 的值是多少？（1分）

(2) 该数据库事务隔离级别最高是什么？请简要给出判断依据。并简述该调度会引发的问题。（4分）

- (3) 假定目前数据库的隔离级别是可重复读，请回答表中的 X、Y、Z 值分别是多少？该隔离级别会有什么问题？如果想要让事务达到串行化，应该使用什么方法？(5 分)

时间	事务 T1	事务 T2	事务 T3
t1	FIND A (100)		
t2		FIND A (100)	FIND A (100)
t3	A := A - 30		
t4			
t5	*COMMIT*	A := A + 50	
t6		FIND A (X)	FIND A (Y)
t7			
t8		*ROLLBACK*	A := A + 70
t9			FIND A (Z)
t10			*COMMIT*