

1. 已知 2 个关系  $R(A, B)$  和  $S(B, C)$ ，其主键分别为  $R.A$  和  $S.B$ 。 $R$  有 20000 个元组， $S$  有 60000 个元组，一块中可以容纳 20 个  $R$  元组或 30 个  $S$  元组。设 2 个关系均采用聚簇存储，且每个关系中的元组均已按照其主键值递增排序。现在要执行自然连接操作  $R \bowtie S$ 。设缓冲区中可用内存页数为  $M = 41$ 。回答下列问题：

(1) 采用嵌套循环连接算法执行  $R \bowtie S$  分别需要进行多少次 I/O？给出具体分析过程。

(2) 采用归并连接算法执行  $R \bowtie S$  分别需要进行多少次 I/O？给出具体分析过程。

(3) 设  $R.B$  是关系  $R$  的外键，参照  $S.B$ 。如果  $R \bowtie S$  的结果中元组的平均大小是  $R$  中元组平均大小的 1.2 倍， $R \bowtie S$  的结果中元组的平均大小是  $S$  中元组平均大小的 2 倍，那么在外存中存储  $R \bowtie S$  的结果需要占用多少个块（页）？给出具体分析过程。

2. 设关系  $R(X, Y)$  和  $S(Y, Z)$ ， $R$  共有 1000 个元组， $S$  共有 1500 个元组，每个块中可容纳 20 个  $R$  元组或 30 个  $S$  元组。 $S$  中  $Y$  不同值的个数为 20。

(1) 若在  $S.Y$  上建有聚簇索引，估计  $R$  和  $S$  基于索引连接的 IO 代价。

(2) 若在  $S.Y$  上建有非聚簇索引，估计  $R$  和  $S$  基于索引连接的 IO 代价。

3. 设教学管理数据库有如下 3 个关系模式：

$S(S\#, SNAME, AGE, SEX)$

$C(C\#, CNAME, TEACHER)$

$SC(S\#, C\#, GRADE)$

其中  $S$  为学生信息表、 $SC$  为选课表、 $C$  为课程信息表； $S\#$ 、 $C\#$  分别为  $S$ 、 $C$

表的主码， $(S\#, C\#)$  是  $SC$  表的主码，也分别是参照  $S$ 、 $C$  表的外码

用户有一查询语句：

Select SNAME

From S, SC, C

Where  $SC.S\#=S.S\#$  and  $SC.C\#=C.C\#$  and CNAME= “数据库”

检索选学“数据库”课程的学生的姓名。

(1) 写出以上 SQL 语句所对应的关系代数表达式。

(2)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。

(3)设 SC 表有 10000 条元组，C 表有 50 条元组，S 表中有 1000 条元组，SC 中满足选修数据库课程的元组数为 150，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

4. 给定以下关系模式，

```
Rabbit (rid, name, color)
Carrot (cid, weight, type)
Allocation (rid, cid, date)
```

(1) 考虑以下的 SQL 查询语句，绘制其查询计划树。

```
SELECT C.type
FROM Rabbit R, Carrot C, Allocation A
WHERE R.rid = A.rid
AND C.cid = A.cid
AND R.name = 'Tutu'
AND C.weight > 100
AND C.weight < 500
```

(2) 假设在 Rabbit.name 和 Allocation.rid 上建有索引，绘制优化后的查询计划树。

5. 已知一个关系数据库的模式如下：

关系 B(bno, bname, author)为图书表，其中 bno 为书号，bname 为书名，author 为作者；

关系 S(sno, sname, dept)为学生表，其中 sno 为学号，sname 为姓名，dept 为学生所在系；

关系 L(sno, bno, date)为借书表，其中 sno 为学号，bno 为书号，date 为借书时间。

回答下列问题：

(1) 绘制下面的 SQL 查询语句的逻辑查询计划树。

```
SELECT bname FROM B NATURAL JOIN S NATURAL JOIN L
WHERE date = '2021-06-04' AND college = 'CS';
```

(2) 使用启发式查询优化方法对上面的逻辑查询计划树进行优化，绘制优化后得到的逻辑查询计划树，具体说明你进行这些优化的理由。

6. 设  $ri(X)$  与  $wi(X)$  分别表示事物  $T_i$  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可串行化的，为什么？

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  $S'$ ：

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

7. 已知下面两个事务：

```
T1:
read(A);
read(B);
if A > B then B := A;

T2:
read(B);
read(A);
if B < 0 then A := B * B;
```

回答下列问题：

- (1) 简述两段锁协议（简称 2PL），并给事务  $T_1$  和  $T_2$  添加加锁和解锁指令，使其遵从 2PL 协议。
- (2) 按照你添加的加锁和解锁指令，给出 2PL 下对  $T_1$  和  $T_2$  的一个正确的调度。
- (3) 什么情况下， $T_1$  和  $T_2$  会发生死锁？请给出一个会发生死锁的调度。
- (4) 在数据库的实际应用中经常会遇到的与锁相关的异常情况，请简述一下常见的死锁检测方法

8. 已知某数据库采用即时更新方法（undo-redo 方法）记录 WAL 日志。设故障发生时 WAL 日志文件内容如下：

```

<T1, begin>
<T1, A, 114, 114514>
<T2, begin>
<T1, B, "hit", "hits">
<T1, commit>
<T3, begin>
<T3, B, "hits", "hitsdb">
<T2, A, 114514, 1919810>
    
```

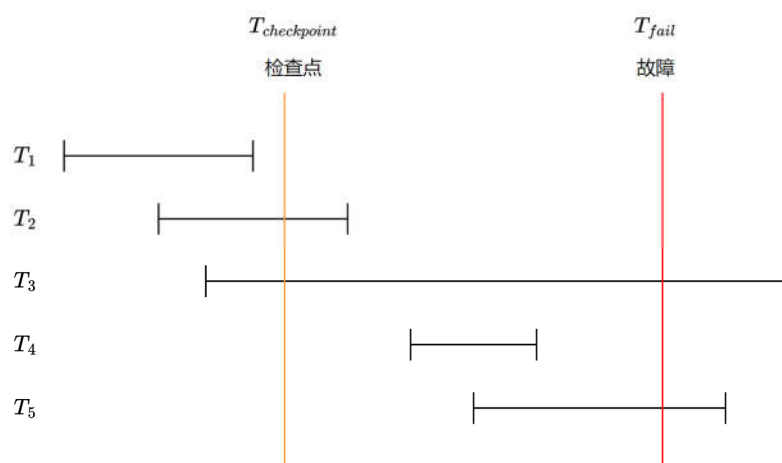
当系统重启后，DBMS 基于该 WAL 日志文件进行故障恢复。回答下列问题：

(1) 该数据库系统的日志恢复策略为 Undo/Redo 型，那么其对应的缓冲区处理策略是什么？该策略的每一项的具体内容都有什么？

(2) 当 DBMS 进行故障恢复时，需要对那些事务进行 undo？对那些事务进行 redo？给出具

(3) 当故障恢复完成时，对象 A 和 B 的值分别是什么？描述故障恢复的具体过程。

9. 使用检查点的数据库恢复系统将根据事务的状态和检查点的关系采取相对应的恢复策略，现在有事务 $T_1$ - $T_5$ 其执行过程如下图所示（线段左端和右端分别表示事务开始与提交），其执行过程中数据库系统发生如图所示的故障，请回答下列问题



(1) 请问在故障恢复时事务 $T_1$ - $T_5$ 那些需要撤销，那些需要重做，那些不需要操作？

(2)事务 $T_6$ - $T_8$  的日志文件如下图所示,< $T_i$ , begin>表示事务 $T_i$ 开始执行, < $T_i$ , commit>表示事务 $T_i$ 提交, < $T_i$ , D,  $V_1$ ,  $V_2$ >表示事务 $T_i$  将数据项 D 的值由 $V_1$ 修改为 $V_2$ , <crash>表示数据库发生故障

```
<T6, begin>
<T6, X, 100, 1>
<T7, begin>
<T7, X, 1,3>
<T8, begin>
<T7, Y, 50, 6>
<T8, Y, 6, 8>
<T8, Z, 10, 9>
<checkpoint>
<T6,commit>
<T8,Z,9,10>
<crash>
```

数据库系统发生故障时, 请给出恢复子系统时需要 undo 的事务列表和需要 redo 的事务列表

(3)请简述事务 $T_6$ - $T_8$  在系统故障后, 基于检查点的故障恢复过程