

《数据库系统原理》期末考试（回忆版）

考试方式：闭卷

考试日期：2025-06-23

考试时长：150 分钟

一、(13 分)

设一个关系数据库中有 R, S 两个关系，定义视图 V_RS 为这两个关系的自然连接。

- NSM 和 DSM 这两种存储模型，哪种更适合 OLAP 型应用？另一种存储模型更适合什么类型的应用？
- 为优化效率，该数据库采用一种模仿某种索引的记录存储方式，可以实现预连接关系 R, S 的效果。这种存储方式是什么？优化后，查询视图 V_RS 的应用程序是否需要修改？这体现了数据库系统的什么特点？
- 将关系 S 无损连接分解为关系 S1, S2。操作后，查询视图 V_RS 的应用程序是否需要修改？这体现了数据库系统的什么特点？
- 数据库的查询处理模型有哪三种？哪一种的内存占用最少？

二、(20 分)

1. 关系代数 (12 分)

假设有以下三个关系 R, S, T:

	R				S				T	
A	B	C		B	C	D		B	C	D
甲	10	红		5	红	80		5	红	80
乙	10	红		10	红	85		10	红	85
丙	20	黄		20	黄	80		20	黄	85
丁	30	兰		30	兰	88		30	兰	88

计算以下关系代数表达式的结果：

- $\Pi_{B,C}(\sigma_{D>80}(S))$ (6 分)
- $(R \bowtie S) \cap (R \bowtie T)$ (6 分)

2. 关系模式分析 (8 分)

给定关系模式 $R(A, B, C, D, E, G)$ 和其上的函数依赖集

$F = \{CG \rightarrow B, AD \rightarrow C, AD \rightarrow E, BCE \rightarrow G, AC \rightarrow B\}$ 。

- 函数依赖 $AD \rightarrow B$ 是否为 F 所逻辑蕴涵？给出证明。(4 分)
- 对该关系模式投影分解，其中一个子关系为 $R_1(B, C, E, G)$ 。判断 R_1 最高属于第几范式，并给出理由。(4 分)

三、 (22 分)

假设有一个关税数据库，包含以下关系表：

- 报关记录表 REPORT(PID, CID, TPID, PRICE, QTY)
PID 为商品编号，CID 为国家编号，TPID 为商品类型编号，PRICE 为商品单价，QTY 为商品数量。
- 税率表 TAX(CID, TPID, RATE)
对于来自编号为 CID 的国家的 TPID 类型商品，税率为 RATE。商品总价乘以税率即为税款。
- 免税商品表 WHITELIST(CID, PID)
对来自编号为 CID 的国家的编号为 PID 的商品，不征收关税。

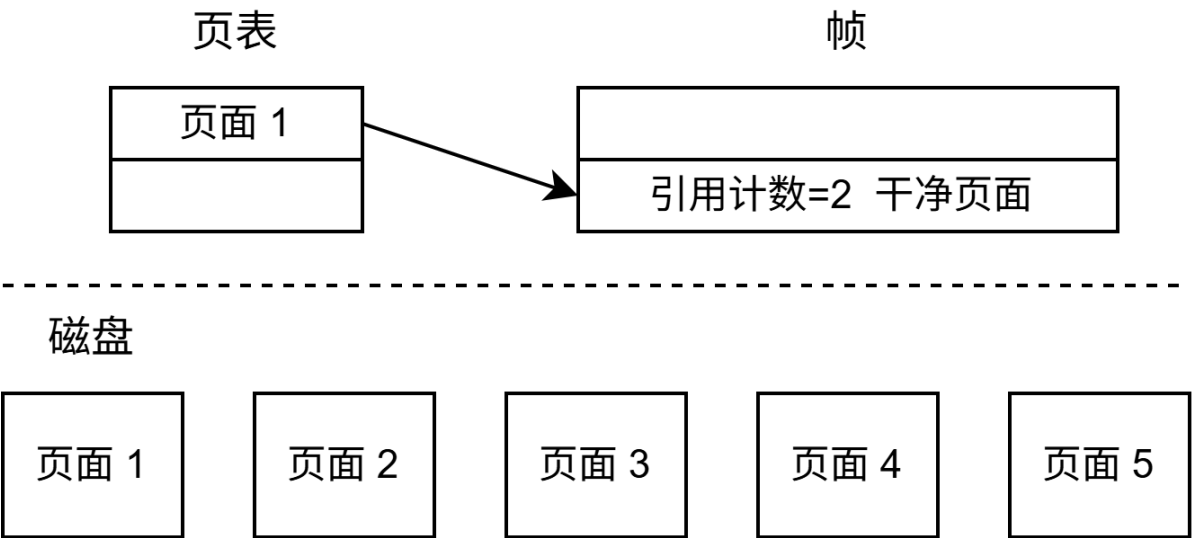
使用单条 SQL 语句，不使用公共表表达式完成以下任务：

1. 查询国家 C1 的所有免税商品编号和单价，按商品编号降序排序。(6 分)
2. 按国家分组，计算每个国家的编号以 A01 开头的商品的总税额。(8 分)
3. 授予用户 U1 修改 TAX 表中的 RATE 列的权限，该权限可以转授。(4 分)
4. 添加一个完整性约束，限制国家 01 的 02 类商品税率不得低于 20%。(4 分)

四、 (10 分)

1. 缓冲区管理 (5 分)

一个数据库在磁盘上占据 5 个页面，缓冲区包含 2 个帧。某一时刻，系统的状态如下：



接下来需要按顺序执行以下两个操作：

1. 事务 T1 更新页面 2 中的元组，随后提交
2. 事务 T2 查询页面 3 中的元组

简要描述执行过程中内外存之间的数据交换，以及恢复子系统可能需要做的记录。

2. 查询优化 (5 分)

对于以下 SQL 查询，画出经过代数优化的语法树。可以合并在一起执行的一组操作用虚线框起。注意尽可能执行能够优化查询效率的投影运算。

```
1 SELECT R.RID, R.RNAME
2 FROM R, RS, S
3 WHERE R.RID > 10 AND R.RID = RS.SID AND RS.SID = S.SID AND S.SID = 99;
```

五、 (10 分)

假设数据库系统发生断电故障，故障前瞬间状态如下，系统重启后内存中的信息全部丢失且不可恢复。A, B, C 的初始值均为 0。

重新开始文件

检查点 1	日志序号 8
-------	--------

日志文件

1	T1: 开始	9	T1: 提交
2	T1: 写A, A=11	10	T4: 开始
3	T2: 开始	11	T5: 开始
4	T2: 写B, B=9	12	T4: 写A, A=5
5	T2: 提交	13	T4: 提交
6	T3: 开始	14	T5: 写A, A=3
7	T3: 写B, B=7	15	检查点 2
8	检查点 1		

内存中的日志

16	T5: 提交
17	T6: 开始
18	T6: 写C, C=1
19	T6: 提交

1. 分析断电后磁盘上 A, B, C 的取值情况；给出恢复完毕后 A, B, C 的值。(5 分)
2. 描述恢复时 UNDO 和 REDO 队列的变化过程。(5 分)

六、 (10 分)

给定以下调度 $S: R_1(X) R_2(Y) W_2(X = Y \times 3) W_1(Y = X + 8) W_2(X) W_3(X = 33)$ ，其中 $R_1(X)$ 表示事务 T1 读取 X 的值， $W_1(Y = X + 8)$ 表示事务 T1 计算其先前读取到的 X 的值 +8，并将结果写入 Y。其余各项类似。X 的初始值为 3，Y 的初始值为 5。

1. 调度 S 是否冲突可串行化？是否可串行化？(3 分)
2. 如果去掉事务 T3，并发执行 T1 和 T2，有哪些可能的正确结果？(3 分)
3. 若对这三个事务 T1, T2, T3 采用 3 级封锁协议，以表格形式给出一种可能的并发调度，并写出最终结果。(4 分)

七、 (15 分)

研究生招生网站需要建立一个数据库。考生可以关注多所学校；一所学校可以开设多个专业；不同学校可以开设相同专业，相同的专业有国家统一的专业代码和专业名；考生只能报名一所学校的一个专业。系统包含的属性有：学校编号、学校名、学校在各个专业的招生人数、专业代码、专业名、考生编号、考生姓名、考生电话、考生报名时选择的学位类型。

1. 根据以上需求画出 E-R 图，注明属性和联系类型。(9 分)
2. 将上一题中的 E-R 图转换为关系模式，并标出主码和外码。(6 分)