

非加粗答案在 PPT 上

Homework #1 (3.13)

(PPT 有) 1、给定下面的关系：图书（图书号，书名，作者，单价，库存量），读者（读者号，姓名，工作单位，地址），借阅（图书号，读者号，借期，还期，备注）注：还期为 NULL 表示该书未还。

使用关系代数表达式实现下列 1—3 小题：

(1) 检索读者 Rose 的工作单位和地址； (2) 检索读者 Rose 所借阅读书（包括已还和为还图书）的图书名和借期； 3) 检索未借阅图书的读者姓名

用 SQL 语言完成 4—8 小题： (4) 检索 Ullman 所写的书的书名和单价 (5) 检索读者“李林”借阅未还的图书的图书号和书名； (6) 检索借阅图书数目超过 3 本的读者姓名； (7) 检索没有借阅读者“李林”所借的任何一本书的读者姓名和读者号； (8) 检索书名中包含“Oracle”的图书书名及图书号。

2、给定关系模式 $R(A,B)$ 、 $S(B,C)$ 和 $T(C,D)$ ，已知有下面的关系代数表达式，其中 p 是涉及属性 $R.A$ 的谓词， q 是涉及 $R.B$ 的谓词， m 是涉及 $S.C$ 的谓词。请写出与此关系代数表达式对应的 SQL 查询语句：

$$\pi (D) [\sigma (p \wedge q \wedge m) (R \bowtie S) \bowtie T]$$

Select D from (select R.*, S.C from R,S Where R.B=S.B and p and q and m) , S,T where S.C=T.C

3、已知有关系模式 $R(A,B,C,D,E)$ ， R 上的一个函数依赖集 $F=\{A \rightarrow BC, B \rightarrow CE, A \rightarrow B, AB \rightarrow C, AC \rightarrow DE\}$ 。

(1) 求 R 上 F 的一个最小函数依赖集（要求写出求解过程）；

1 右边写出单属性并去除重复

$F=\{A \rightarrow B, A \rightarrow C, B \rightarrow C, B \rightarrow E, AB \rightarrow C, AC \rightarrow D, AC \rightarrow E\}$ 。

2 消除左部冗余

$F=\{A \rightarrow B, A \rightarrow C, B \rightarrow C, B \rightarrow E, A \rightarrow D, A \rightarrow E\}$

3 消除冗余函数依赖

$F=\{A \rightarrow B, B \rightarrow C, B \rightarrow E, A \rightarrow D\}$

(2) 求 R 的候选码，并给出证明。

候选码为 A。因为 $A \rightarrow B, A \rightarrow C, A \rightarrow D, A \rightarrow E$ ，并且不存在 A 的真子集 Y 使得 $Y \rightarrow U$ 成立。

Homework #2 (3.20)

1、现有如下关系模式：R(A, B, C, D, E, F, G)，R 上存在的函数依赖有： $AB \rightarrow E, A \rightarrow B, B \rightarrow C, C \rightarrow D$

(1) 该关系模式满足第几范式吗？为什么？

(2) 如果将关系模式 R 分解为： $R_1(A, B, E), R_2(B, C, D), R_3(A, F, G)$ ，该数据库模式最高满足第几范式？

(3) 请将关系模式 R 无损连接并且保持函数依赖地分解到 3NF，要求给出具体步骤。PPT 3-1

(4) 请将关系模式 R 无损连接地分解到 BCNF，要求给出步骤 PPT3-1

(C.J. Date) 如果关系模式 R 的所有不平凡的功能完全函数依赖的决定因素（左边的属性集）都是候选码，则 $R \in BCNF$

(萨 & 王) 关系模式 $R \in 1NF$ ，若 R 中的任一函数依赖 $X \rightarrow Y$ 且 $Y \not\subseteq X$ 时 X 必包含候选码，则 $R \in BCNF$

- **3NF**：不允许非主属性到非码的 FD，但允许主属性到其它属性的 FD
- **BCNF**：不允许主属性、非主属性到非码的 FD

■ $R(S\#, C\#, G, TN, D),$
 $F = \{\{S\#, C\#\} \rightarrow G, C\# \rightarrow TN, TN \rightarrow D\}$

1. $p = \{R\};$
2. $TN \rightarrow D$ 不满足 BCNF 定义，分解 R
 $p = \{R_1(S\#, C\#, G, TN), R_2(TN, D)\}$
3. R_1 中 $C\# \rightarrow TN$ 不满足 BCNF，分解 R_1 为 R_3 和 R_4
 $p = \{R_3(S\#, C\#, G), R_4(C\#, TN), R_2(TN, D)\}$
4. p 中各模式均满足 BCNF，结束

Homework #3 (4.3)

1、假设某磁盘具有以下特性：

- (1) 有 8 个盘面和 8192 个柱面
- (2) 盘面直径为 3.5 英寸，其中内圈不存储数据，内圈直径为 1.5 英寸
- (3) 每磁道平均有 256 个扇区，每个扇区 512 字节
- (4) 每个磁道 10% 被用于间隙
- (5) 磁盘转速为 7200 RPM
- (6) 磁头启动到停止需要 1ms，每移动 500 个柱面另加 1ms

回答下列问题：

- (1) 磁盘容量是多少？

$$2^3 * 2^{13} * 2^8 * 2^9 = 2^{33} = 8\text{GB}$$

- (2) 如果所有的磁道拥有相同的扇区数，那么最内圈的磁道的位密度是多少？

$$256 * 512 * 8 / (1.5 \pi * 0.9) = 247238.8 \text{ bpi}$$

- (3) 如果一个块是 8KB，那么一个块的传输时间是多少？

$$8\text{KB} / 0.5\text{KB} = 16 \text{ 个扇区}, 15 \text{ 间隙} (512 \text{ 字节} = 0.5\text{KB} = 512\text{byte})、$$

$$7200\text{RPM} = 7200 / 60 = 120\text{RPS},$$

$$\text{转一圈时间 } 1000\text{ms} / 120 = 8.3\text{ms}$$

$$T = 16 * 8.3\text{ms} / 256 - 8.3\text{ms} / 256 * 10\% = 0.52\text{ms}$$

- (4) 平均寻道时间是多少？

$$1 + 8192 / 500 / 3 = 6.5\text{ms}$$

- (5) 平均旋转等待时间是多少？

$$(60 * 1000 / 7200) / 2 = 4.17\text{ms}$$

2、假设某块磁盘的参数如下：容量为 36.7GB，传输速率为 45MB/s，旋转一圈的时间为 4ms，平均寻道时间为 5ms，最小寻道时间为 0.65ms（指磁头寻道到相邻磁道的时间），一个磁道大小为 180KB。如果磁盘块大小为 4KB，请回答下面问题（所有结果均四舍五入保留小数点后两位）：

（1） 随机读取 1000 个磁盘块需要多少时间（ms）？

$$T=1000*(t_{\text{平均寻道}}+t_{\text{平均旋转}}/2+t_{\text{传输}})$$

$$45\text{MB/S}=45\text{KB/ms}$$

$$T_{\text{传输}}=4\text{kb}/45=0.09\text{ms}$$

$$T_r=1000*(5+2+0.09)=7090\text{ms}$$

（2） 假定（1）中的 1000 个磁盘块在单个磁道上连续存储，并且所有磁盘块存储在相邻的磁道上，此时读取这 1000 个磁盘块需要多少时间（ms）？

$$T_s=t_{\text{平均寻道}}+t_{\text{旋转}}/2+1000*t_{\text{传输}}+k*t_{\text{最小寻道}}$$

$$4*1000/180(\text{向上取整})=23, k=22$$

$$T_s=5+2+1000*0.09+22*0.65=111.3\text{ms}$$

HW4

2.5.1 假设一条记录有如下顺序的字段：一个长度为 23 的字符串，一个字符串，一个 2 字节整数，一个 SQL 日期，一个 SQL 时间（无小数点）。字段可在任何字节处开始？

2.6.9 假设我们有 4096 字节块，块中存储 200 字节长的记录。块首部由一个偏移量表组成，它使用字节长的记录。块首部由一个偏移量表组成，它使用 2 字节长指针指向块内记录。通常，每天向每块插入两条记录，删除一条记录。删除记录必须使用一个“删除标记”代替它的指针，因为可能会有悬挂指针指向它。更明确地说，假设任何一天删除记录总发生在插入之前。如果刚开始时块是空的，多少天之后，不再有插入记录的空间？

2.7.1 一个病人记录包含以下定长字段： 病人的出生日期， 社会保险号码， 病人 ID， 每一个字段都是 9 字节长。它还有下列变长字段： 字节长。它还有下列变长字段： 姓名，住址和病史。如果记录内一个指针需要。如果记录内一个指针需要 8 字节，记录长度是一个 2 字节整数，不包含变长字段空间，这条记录需要多少字节？你可以假设不需要对字节进行对齐。

HW5

5.4.1(PPT 第七章)

5.4.8 习题

习题 5.4.1 下面是 4 个关系 W、X、Y、Z 的关键统计值：

| $W(a, b)$ | $X(b, c)$ | $Y(c, d)$ | $Z(d, e)$ |
|----------------|-----------------|----------------|----------------|
| $T(W) = 400$ | $T(X) = 300$ | $T(Y) = 200$ | $T(Z) = 100$ |
| $V(W, a) = 50$ | $V(X, b) = 60$ | $V(Y, c) = 50$ | $V(Z, d) = 10$ |
| $V(W, b) = 40$ | $V(X, c) = 100$ | $V(Y, d) = 20$ | $V(Z, e) = 50$ |

估计下列表达式结果关系的大小：

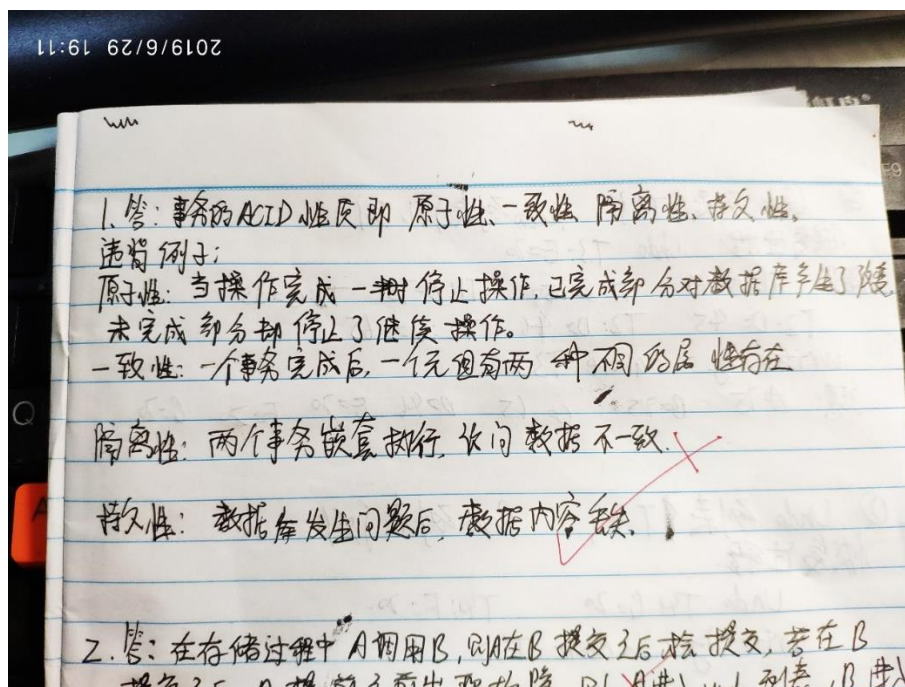
- | | | |
|---------------------------------------|---------------------------------------|------------------------------|
| a) $W \bowtie X \bowtie Y \bowtie Z$ | b) $\sigma_{a=10}(W)$ | c) $\sigma_{c=20}(Y)$ |
| d) $\sigma_{c=20}(Y) \bowtie Z$ | e) $W \times Y$ | f) $\sigma_{d>10}(Z)$ |
| g) $\sigma_{a=1 \text{ AND } b=2}(W)$ | h) $\sigma_{a=1 \text{ AND } b>2}(W)$ | i) $X \bowtie_{X.c < Y.c} Y$ |

4.3.5 如果 R 和 S 都是非聚集的，似乎嵌套循环连将需要大约 $T(R)T(S)/M$ 次磁盘 I/O 时间。你怎样做才能明显好于这个代价？

4.4.5 假设这节中所描述算法的第二趟不需要所有的 M 个缓冲区，因为子表数小于个缓冲区，因为子表数小于 M。我们怎样通过使用额外的缓冲区来节省磁盘。我们怎样通过使用额外的缓冲区来节省磁盘 I/O？

Homework #6

1. 什么是事务的 ACID 性质？请给出违背事务 ACID 性质的具体例子，每个性质举一个例子。



2. 存储过程中允许使用事务编程，并且存储过程可以调用其它存储过程。假设一个存储过程 A 调用了另一个存储过程 B，如果 A 和 B 都使用了事务编程，请分析一下这种情况下会出现什么问题？

此时如果事物 B 提交, A 调用 B 返回后, 事物 A 选择 roll back, 此时由于 B 已经提交, 持久性已经生效, A 的 roll back 已经不能将事务 B 撤销, 从而导致事物 A 的原子性和一致性被破坏。

3. 下面是一个数据库系统开始运行后的日志记录, 该数据库系统支持检查点。

- 1) <T1, Begin Transaction>
- 2) <T1, A, 49, 20>
- 3) <T2, Begin Transaction>
- 4) <T1, B, 250, 20>
- 5) <T1, A, 75, 49>
- 6) <T2, C, 35, 20>
- 7) <T2, D, 45, 20>
- 8) <T1, Commit Transaction>
- 9) <T3, Begin Transaction>
- 10) <T3, E, 55, 20>
- ①
- 11) <T2, D, 46, 45>
- 12) <T2, C, 65, 35>
- 13) <T2, Commit Transaction>
- ②
- 14) <T3, Commit Transaction>
- 15) <CHECKPOINT>
- 16) <T4, Begin Transaction>
- 17) <T4, F, 100, 20>
- 18) <T4, G, 111, 20>
- ③
- 19) <T4, F, 150, 100>
- 20) <T4, Commit Transaction>

设日志修改记录的格式为 <Tid, Variable, New value, Old value>, 请给出对于题中所示①、②、③三种故障情形下, 数据库系统恢复的过程以及数据元素 A, B, C, D, E, F 和 G 在执行了恢复过程后的值。

3. 解: ① Undo { T2, T3 } Redo { T1 }
 恢复过程: undo: T3: E=20 T2: D=20 T2: C=20
 Redo: T1: A=49 T1: B=250 T1: A=75
 write log < Abort, T2 >
 < Abort, T3 >
 结果: A=75 B=250 C=20 D=20 E=20 F=20 G=20

1、基于Undo/Redo日志的恢复

- 正向扫描日志，将 $\langle \text{commit} \rangle$ 的事务放入Redo列表中，将没有结束的事务放入Undo列表
- 反向扫描日志，对于 $\langle T, x, v, w \rangle$ ，若T在Undo列表中，则
 - Write(x,v); Output(x)
- 正向扫描日志，对于 $\langle T, x, v, w \rangle$ ，若T在Redo列表中，则
 - Write(x,w); Output(x)
- 对于Undo列表中的T，写入 $\langle \text{abort}, T \rangle$

② Undo 列表 {T3} Redo 列表 {T1, T2}

恢复过程: Undo T3: E=20

Redo T1: A=49 T1: B=250 T1: A=75 T2: C=35

T2: D=45 T2: D=46 T2: C=65

write log: $\langle \text{abort T3} \rangle$

结果: A=75 B=250 C=65 D=46 E=20 F=20 G=20

① Undo 列表 {T4} Redo 列表 { }.

恢复过程:

Undo T4: G=20 T4: F=20.

write log: $\langle \text{Abort T4} \rangle$

结果 A=75 B=250 C=65 D=46 E=55
F=20 G=20.