

高级数据库系统习题解答 (2)

第五次作业

- 1.1（书本概念，略）
- 1.3（书本概念，略）
- 1.5（书本概念，略）

第六次作业

- 已知有如下两种段分配:

A> R1在Site1, R2在Site2, R3在Site3.

B> R1和R2在Site1, R2和R3在Site3.

另已知有如下应用(所有应用的频率相同)

A1: 在Site1上发出, 读5个 R1记录, 5个 R2记录

A2: 在Site3上发出, 读5个R3记录, 5个R2记录

A3: 在Site2上发出, 读10个R2记录.

问: 1. 如果以本地应用为主要设计目标, 那个分配较优?

2. 假定A3改为要修改10个R2记录, 并仍以本地应用为其设计目标, 则那个分配方案较优?

第六次作业

- 解：首先完善模型如下：

假定：B方案中Site1上的R2和Site3上的R2是完全相同的，即冗余；本地读写的代价可忽略不计；所有异地读操作的单位记录代价均相等，记为1；所有异地写操作的单位记录代价均相等，记为1。

	1 A方案	B方案	2 A方案	B方案
A1代价	5	0	5	0
A2代价	5	0	5	0
A3代价	0	10	0	10+10+10

故可知，第1个问题，A、B两种分配方案代价相等。

第2个问题，A分配方案较优。

第六次作业

- 2.2

解：分布式数据库设计的目标除包括集中式数据库设计中的目标外，还包括：

- 分布式数据库的本地性或近地性
- 控制数据的适当冗余
- 工作负荷分布
- 存储的能力和费用

- 2.6

解：分片基本原则：完整性，可重构，不相交。分片方法：水平分片，垂直分片，混合分片

- 2.9（过程略，参见课后习题答案）

第七次作业

- 3.6

解：(1)

R J N S =

A	B	C	D
2	3	5	6
2	3	5	9
1	6	8	3
5	3	5	6
5	3	5	9
2	6	8	3

R J N S J N T =

A	B	C	D	E	I
2	3	5	6	6	9
1	6	8	3	8	9
5	3	5	6	6	9
2	6	8	3	8	9

第七次作业

- 3.6

解: (2)

R S J S =

A	B	C
2	3	5
1	6	8
5	3	5
2	6	8

S S J T =

B	C	D
3	5	6
6	8	3
5	9	6
4	1	6

T S J R 结果为空
R S J T 结果为空

S S J R =

B	C	D
3	5	6
3	5	9
6	8	3

T S J J =

D	E	I
6	6	9
3	8	9

第七次作业

• 3.8

设某公司的雇员关系为employee(name, address, salary, plant-number), 按plant-number水平分片这个关系, 每个片段都有两个副本: 一个副本存放在NewYork站点, 另一个副本存放在工厂的站点。请为在Toronto站点提出的下列查询设计一个好的处理策略。

- (1) 找出Boce厂的所有雇员。
- (2) 找出所有雇员的平均工资。
- (3) 找出在如下每个站点工资最高的雇员姓名: Toronto, Edmonton, Vancouver, Montreal。

第七次作业

- 3.8

解：

- 1)将New York站点上的副本传至Toronto站点；
- 2)在New York站点上求平均工资，传至Toronto站点；
- 3)Toronto, Edmonton, Vancouver, Montreal求最高工资，传至Toronto汇总；

第七次作业

- PPT第13章补充题

假定全局关系和分段模式如下：

全局关系

Doctor (D#, Name, Dept)

Patient (P#, Name, Dept, Treat, D#)

Care (P#, Drug, QTY)

分段模式

Doctor1 = SL_{DEPT = 'Surgery'} (Doctor)

Doctor2 = SL_{DEPT = 'Pediatrics'} (Doctor)

Doctor3 = SL_{DEPT ≠ 'Surgery' and DEPT ≠ 'Pediatrics'} (Doctor)

Patient1 = SL_{DEPT = 'Surgery' and treat = 'intensing'} (Patient)

Patient2 = SL_{DEPT = 'Surgery' and treat ≠ 'intensing'} (Patient)

Patient3 = SL_{DEPT ≠ 'Surgery'} (Patient)

Care1 = Care SJ_{p#=p#} Patient1

Care2 = Care SJ_{p#=p#} Patient2

Care3 = Care SJ_{p#=p#} Patient3

使用变换准则，将下面的全局查询变化为分段查询，并对它们加以简化。当需要时用限定关系代数消除查询中不用的段。

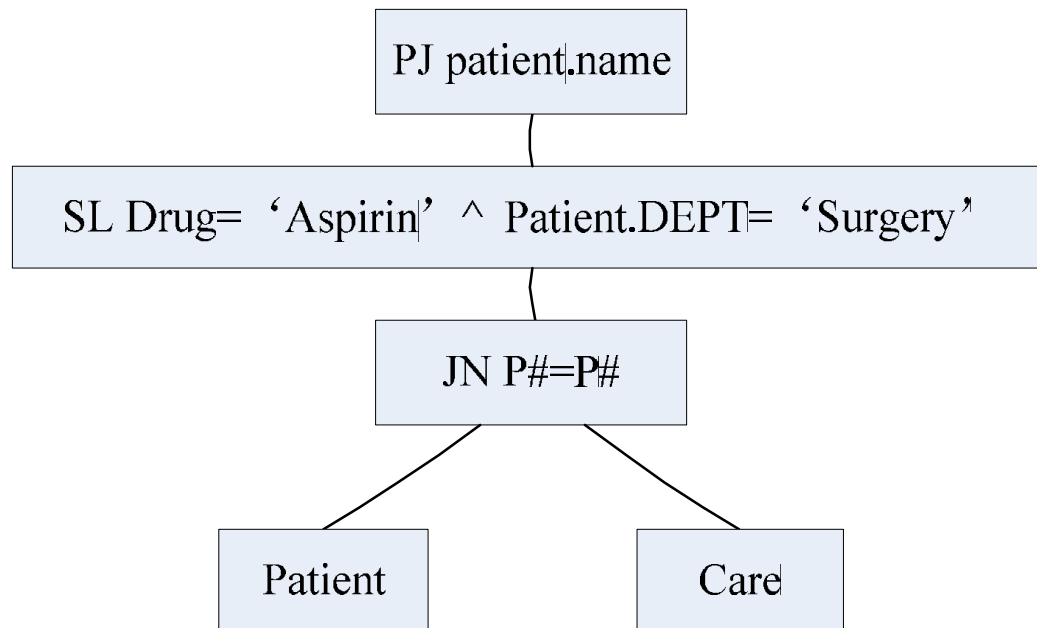
- (a) 列出在Care中使用‘Aspirin’的Patient名字；
- (b) 列出给病人开出接受精细护理的医生名字。

第七次作业

- PPT第13章补充题

解： (a) 列出在Care中使用 ‘Aspirin’ 的Patient 名字

$$\pi_{Patient.Name} \left(\sigma_{Care.Drug='Aspirin' \wedge Patient.DEPT='Surgery'} (Patient \triangleright \triangleleft Care) \right)$$

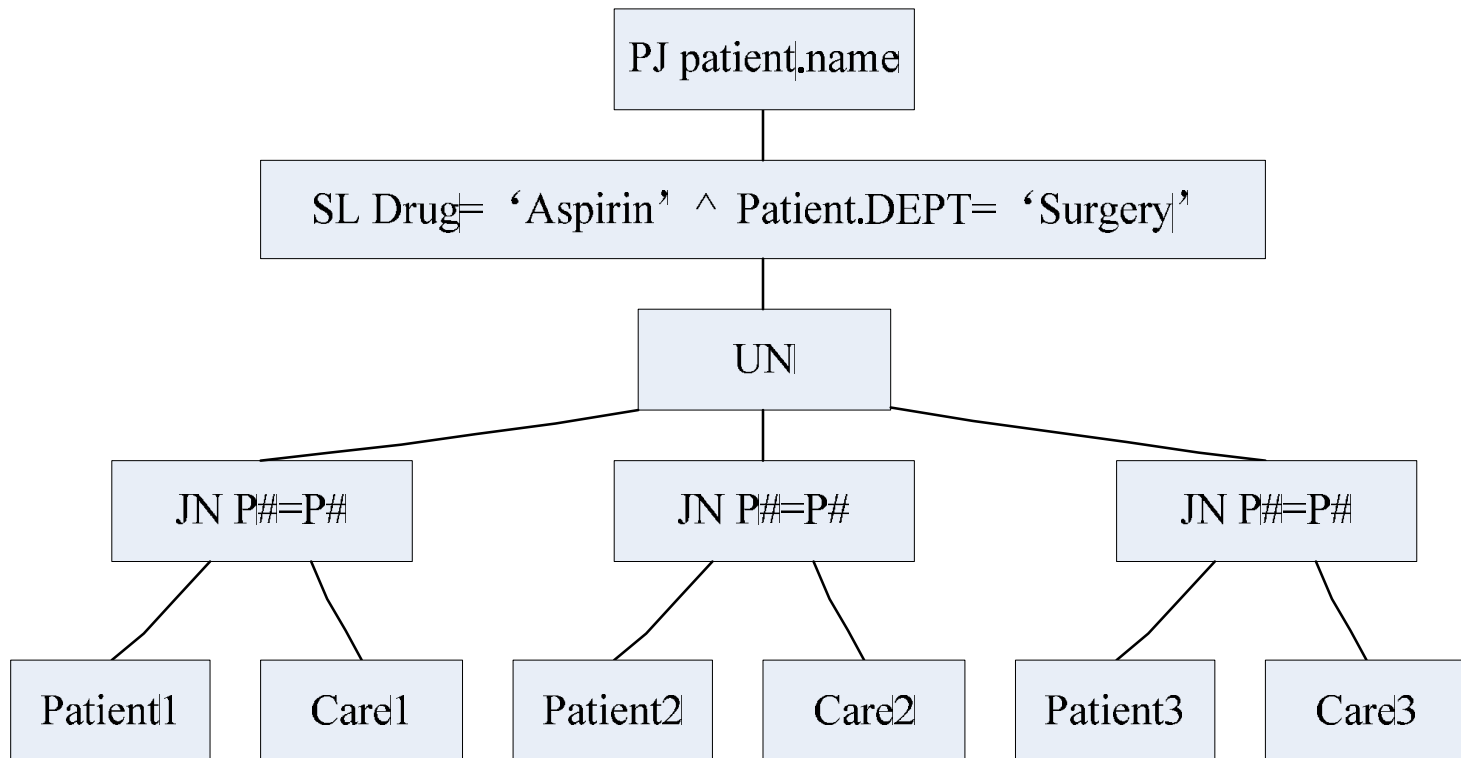


第七次作业

- PPT第13章补充题

解： (a) 列出在Care中使用 ‘Aspirin’ 的Patient 名字

$$\pi_{Patient.Name} \left(\sigma_{Care.Drug='Aspirin' \wedge Patient.DEPT='Surgery'} (Patient \triangleright \triangleleft Care) \right)$$

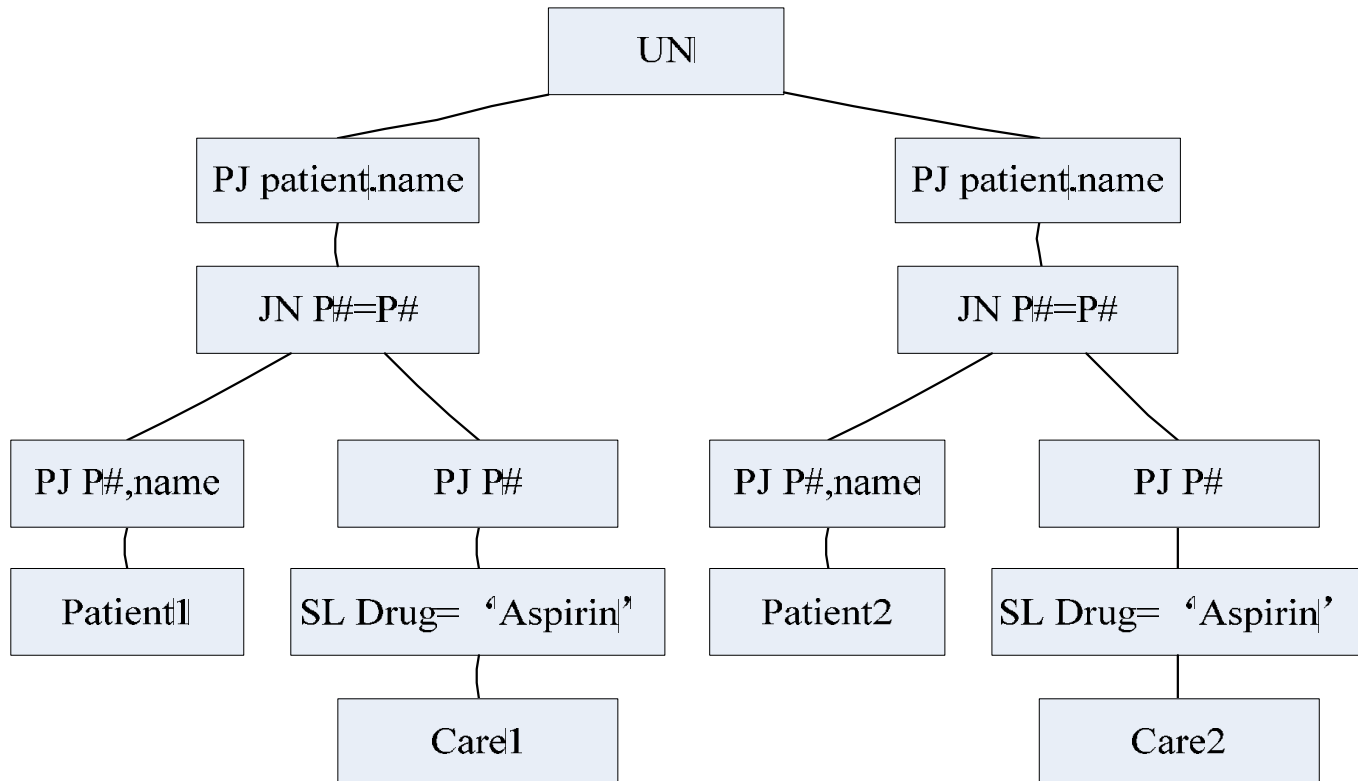


第七次作业

- PPT第13章补充题

解： (a) 列出在Care中使用 ‘Aspirin’ 的Patient 名字

$$\pi_{Patient.Name} \left(\sigma_{Care.Drug='Aspirin' \wedge Patient.DEPT='Surgery'} (Patient \triangleright \triangleleft Care) \right)$$

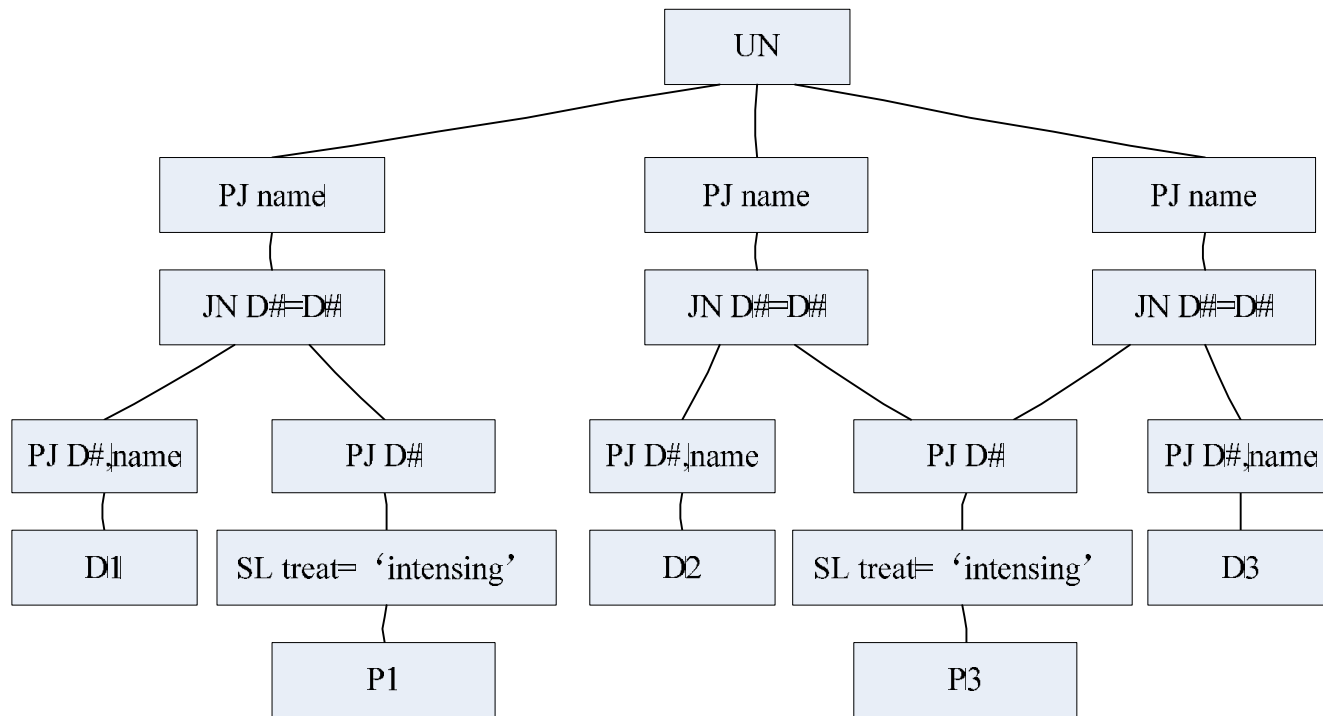


第七次作业

- PPT第13章补充题

解： (b) 列出给病人开出接受精细护理的医生名字

$$\pi_{Doctor.Name} \left(\sigma_{Patient.Treat='Intensing'} \left(Doctor \bowtie_{Doctor.D\#=Patient.D\#} Patient \right) \right)$$



第七次作业

- PPT第14章补充题

在如下R, S的概貌上计算R JN_{A=B} S

Size(R)=50, Card(R)=100, Val(A[R])=50, Size(A)=3

Size(S)=5, Card(S)=50, Val(B[S])=50, Size(B)=3

R SJ_{A=B} S 的选择度 $\rho = 0.2$

S SJ_{A=B} R 的选择度 $\rho = 0.8$

问:

1. 使用 SJ简化程序在R得站点执行JN
2. 使用 SJ简化程序在S得站点执行JN
3. 使用直接连接在R站点执行JN
4. 使用直接连接在S站点执行JN

那种方案较优?

第七次作业

- PPT第14章补充题

解：

$$\begin{aligned} 1. \quad \text{COST1} &= 2C_0 + C_1(\text{Size}(A) \cdot \text{Val}(A[R]) + \text{Size}(S) \cdot \text{Card}(S) \cdot \rho_{S \text{ SJ } R}) \\ &= 2C_0 + C_1(3 \cdot 50 + 5 \cdot 50 \cdot 0.8) = 2C_0 + 350C_1 \end{aligned}$$

$$\begin{aligned} 2. \quad \text{COST2} &= 2C_0 + C_1(\text{Size}(B) \cdot \text{Val}(B[S]) + \text{Size}(R) \cdot \text{Card}(R) \cdot \rho_{R \text{ SJ } S}) \\ &= 2C_0 + C_1(3 \cdot 50 + 50 \cdot 100 \cdot 0.2) = 2C_0 + 1150C_1 \end{aligned}$$

$$\begin{aligned} 3. \quad \text{COST3} &= C_0 + C_1 \cdot \text{Size}(S) \cdot \text{Card}(S) \\ &= C_0 + C_1 \cdot 5 \cdot 50 = C_0 + 250C_1 \end{aligned}$$

$$\begin{aligned} 4. \quad \text{COST4} &= C_0 + C_1 \cdot \text{Size}(R) \cdot \text{Card}(R) \\ &= C_0 + C_1 \cdot 50 \cdot 100 = C_0 + 5000C_1 \end{aligned}$$

所以第三种方案最优。

第七次作业

- 3.7

如果习题中的三个关系R,S,T分别位于三个不同的站点X,Y,Z。若采用基于半连接的算法计算 $R \bowtie S \bowtie T$, 请选择使得传输代价最小的连接执行站点和确定半连接序列。

第七次作业

• 3.7

解：假设每个属性域长度均为1B，考虑所有的半连接

方案	半连接	ρ	得益	费用	传送属性
P1	$R \bowtie S$	$2/3$	$1/3 * 3 * 6$	$2 * 5$	S.B S.C
P2	$S \bowtie R$	$1/2$	$1/2 * 3 * 6$	$2 * 4$	R.B R.C
P3	$S \bowtie T$	$2/3$	$1/3 * 3 * 6$	$1 * 3$	T.D
P4	$T \bowtie S$	$1/2$	$1/2 * 3 * 4$	$1 * 4$	S.D

a) 选择得益最高的P2进行优化，得到新的R,S',T，并对受到影响的方案重新计算得益和费用

第七次作业

- 3.7

解：新的R, S', T如下

R =

A	B	C
2	3	5
5	3	6
1	6	8
3	4	6
5	3	5
2	6	8

S' =

B	C	D
3	5	6
3	5	9
6	8	3

T =

D	E	I
6	6	9
8	7	8
8	5	6
3	8	9

第七次作业

• 3.7

解：对受到影响的方案重新计算得益和费用

方案	半连接	ρ	得益	费用	传送属性
P1	$R \infty S'$	$2/3$	$1/3 * 3 * 6$	$2 * 3$	$S'.B S'.C$
P2	$S \infty R$	$1/2$	$1/2 * 3 * 6$	$2 * 4$	$R.B R.C$
P3	$S' \infty T$	$2/3$	$1/3 * 3 * 3$	$1 * 3$	$T.D$
P4	$T \infty S'$	$1/2$	$1/2 * 3 * 4$	$1 * 3$	$S'.D$

b) 选择得益最高的P4进行优化，得到新的R,S',T'，并对受到影响的方案重新计算得益和费用。

第七次作业

- 3.7

解：新的R, S', T'如下

R =

A	B	C
2	3	5
5	3	6
1	6	8
3	4	6
5	3	5
2	6	8

S' =

B	C	D
3	5	6
3	5	9
6	8	3

T' =

D	E	I
6	6	9
3	8	9

第七次作业

• 3.7

解：对受到影响的方案重新计算得益和费用

方案	半连接	ρ	得益	费用	传送属性
P1	$R \infty S'$	$2/3$	$1/3 * 3 * 6$	$2 * 3$	$S'.B S'.C$
P2	$S \infty R$	$1/2$	$1/2 * 3 * 6$	$2 * 4$	$R.B R.C$
P3	$S' \infty T'$	$2/3$	$1/3 * 3 * 3$	$1 * 2$	$T'.D$
P4	$T \infty S'$	$1/2$	$1/2 * 3 * 4$	$1 * 3$	$S'.D$

c) 选择得益最高的P1进行优化，得到新的 R', S', T' ，并对受到影响的方案重新计算得益和费用。

第七次作业

- 3.7

解：新的R', S', T'如下

R' =

A	B	C
2	3	5
1	6	8
5	3	5
2	6	8

S' =

B	C	D
3	5	6
3	5	9
6	8	3

T' =

D	E	I
6	6	9
3	8	9

第七次作业

• 3.7

解：对受到影响的方案重新计算得益和费用

方案	半连接	ρ	得益	费用	传送属性
P1	$R \infty S'$	$2/3$	$1/3 * 3 * 6$	$2 * 3$	$S'.B S'.C$
P2	$S \infty R$	$1/2$	$1/2 * 3 * 6$	$2 * 4$	$R.B R.C$
P3	$S' \infty T'$	$2/3$	$1/3 * 3 * 3$	$1 * 2$	$T'.D$
P4	$T \infty S'$	$1/2$	$1/2 * 3 * 4$	$1 * 3$	$S'.D$

d) 选择得益最高的P3进行优化，得到X,Y,Z站点上最终的R',S'',T'。

第七次作业

- 3.7

解：X,Y,Z站点上最终的R',S'',T'如下

R' =

A	B	C
2	3	5
1	6	8
5	3	5
2	6	8

S'' =

B	C	D
3	5	6
6	8	3

T' =

D	E	I
6	6	9
3	8	9

第七次作业

- 3.7

解：

所以选择各站点做连接的代价为：

$$X\text{站点代价}=2*3+2*3=12$$

$$Y\text{站点代价}=4*3+2*3=18$$

$$Z\text{站点代价}=4*3+2*3=18$$

故选择X站点作为收集站点代价最低。

第七次作业

• 3.7

解：

由简化过程得知半连接过程为：

1. $S' = S \bowtie R$
2. 将 S' 传送给 T ,做半连接 $T \bowtie S'$ 得到 T'
3. 将 S' 传送给 R ,做半连接 $R \bowtie S'$ 得到 R'
4. 将 T' 传送给 S' ,做半连接 $S' \bowtie T'$ 得到 S''

即：

$$(R \bowtie (S \bowtie R)) \bowtie ((S \bowtie R) \bowtie (T \bowtie (S \bowtie R))) \bowtie (T \bowtie (S \bowtie R))$$

第八次作业

- PPT第15章补充题

假设两个事务T和U的log记录如下所示:

$\langle T, \text{Start} \rangle, \langle T, A, 10 \rangle, \langle U, \text{Start} \rangle, \langle U, B, 20 \rangle, \langle T, C, 30 \rangle, \langle U, D, 40 \rangle, \langle U, \text{Commit} \rangle, \langle T, E, 50 \rangle, \langle T, \text{Commit} \rangle$

如果系统故障时, 磁盘上记录的Log记录如下, 请描述数据库恢复管理器的动作.

- a) $\langle U, \text{Start} \rangle$
- b) $\langle U, \text{Commit} \rangle$
- c) $\langle T, E, 50 \rangle$
- d) $\langle T, \text{Commit} \rangle$

第八次作业

- PPT第15章补充题

解：

(a)undo T, undo U

(b)undo T, redo U

(c)undo T, redo U

(d)redo T, redo U

注：

1. 事务开始了没有提交，须执行undo；
2. 事务开始了并提交了，须执行redo；
3. Undo操作先执行，Redo操作后执行。

第八次作业

- 4.5

日志、档案库和检查点的作用是什么？典型的日志包含哪些内容？为什么要“先写日志”？

解：书本概念，略。

- 4.7

用自己的语言描述两阶段提交过程。

解：书本概念，略。

第九次作业

- 第16章PPT

- 1. 指出下面的冲突等价调度(忽略提交C和夭折A动作)

S1=W2(x),W1(x),R3(x),R1(x),C1,W2(y),R3(z),C3,R2(x),C2

S2=R3(z),R3(y),W2(y),R2(z),W1(x),R3(x),W1(x),R1(x),C1,C2,C3

S3=R3(z),W2(x),W2(y),R1(x),R3(x),R2(z),R3(y),C3,W1(x),C2,C1

S4=R3(z),W2(x),W2(y),C2,W1(x),R1(x),A1,R3(x),R3(z),R3(y),C3

解： 这题答案不唯一

第九次作业

- 第16章PPT

- 2. 上题S1-----S4中, 哪个是可串调度?

解:

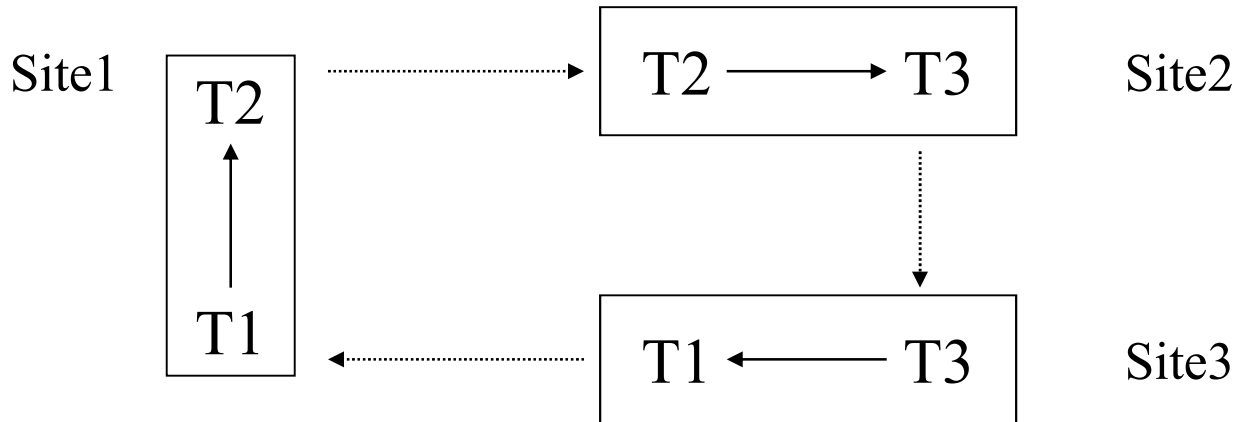
S3和S4为可串调度

S1: T1等待T2, T2等待T1 (W2(x),W1(x),R2(x))

S2: T3等待T1, T1等待T3 (W1(x),R3(x),W1(x))

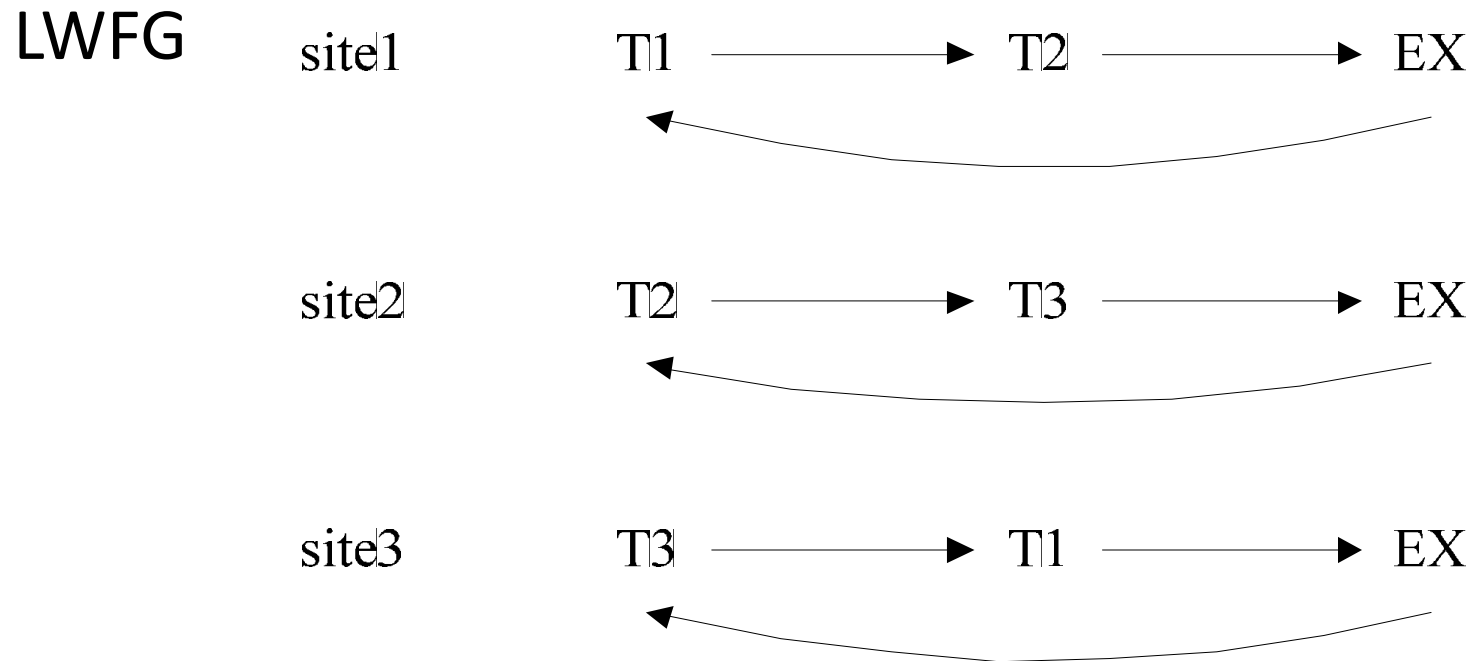
第十次作业

- 第16章PPT
 - 3. 有如下的等待图, 用分布式检测算法检测其是否有死锁(给出检测过程)



第十次作业

- 第16章PPT
 - 3. 有如下的等待图, 用分布式检测算法检测其是否有死锁(给出检测过程)

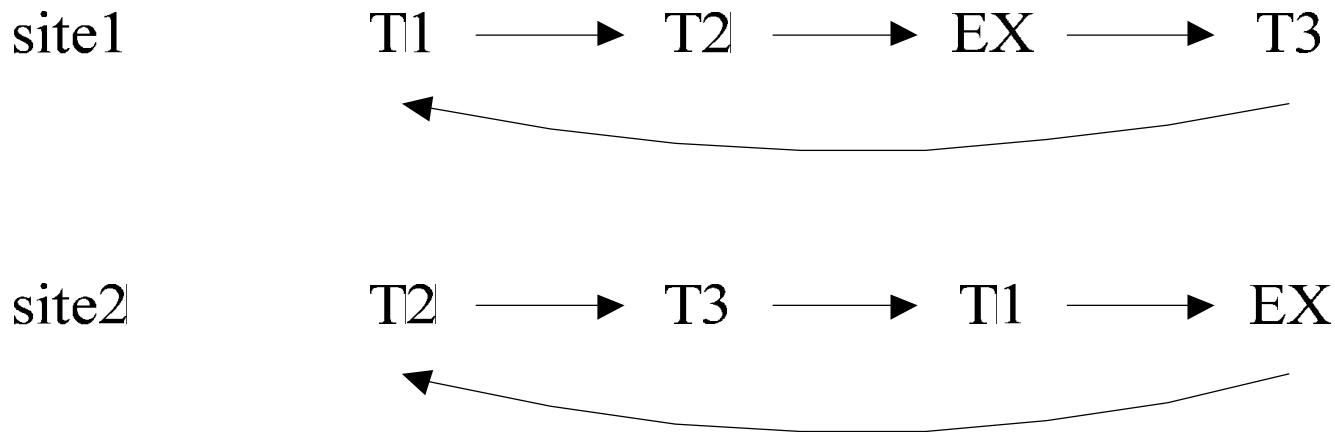


第十次作业

- 第16章PPT

- 3. 有如下的等待图, 用分布式检测算法检测其是否有死锁(给出检测过程)

将site3的LWFG传给site1和site2

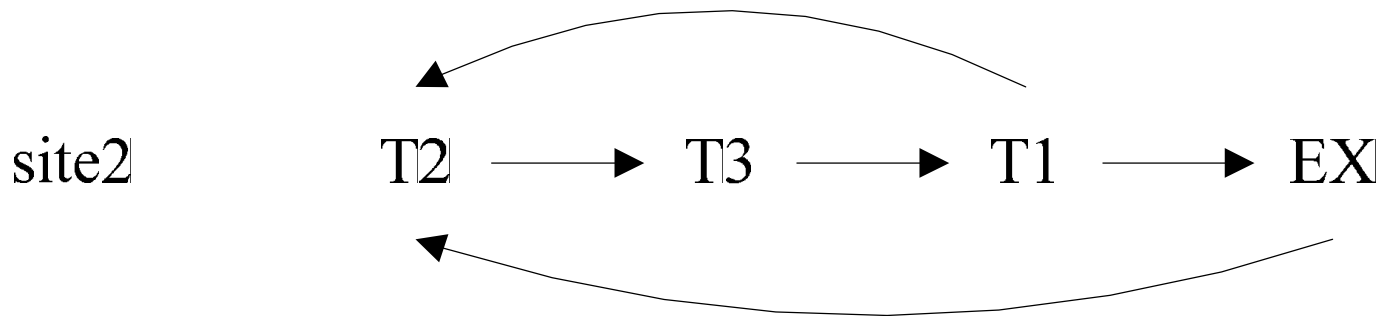


第十次作业

- 第16章PPT

- 3. 有如下的等待图, 用分布式检测算法检测其是否有死锁(给出检测过程)

将site1的信息传给site2, 得到GWFG, 其中存在不含EX的循环, 故存在死锁



第十次作业

- 第16章PPT
 - 4. 已知有事务如例5-1(p.136), 并且T1的TS是20, T2的TS是25重新启动间隔是5, 给出按基本T/O方式执行的过程(调度)

第十次作业

- 第16章PPT

– 4. 解:

T1(20)	T2(25)	T1(30)	RTM(x)	WTM(x)	RTM(y)	WTM(y)
Read(x)			20	0	0	0
	Read(x)		25			
Write(x)						
Abort	Write(x)			25		
	Read(y)	Read(x)	30		25	
	Write(y)	Write(x)		30		25
	Commit	Read(y)			30	
		Write(y)				30
		Commit				

第十次作业

- 6.5: 3PC协议请参见书本内容
- 6.10: 请参考书后答案

关于作业

- 总共十次作业
- 作业提交情况统计
 - <http://home.ustc.edu.cn/~linsh/addb.html>
 - 本PPT可以到上面的网址下载
- 补交作业
 - 电三楼510
 - 电四楼420
 - 截止日期: 2012.01.02

关于实验

- 截止日期
 - 2012.01.02
- 提交方式
 - FTP地址: <ftp://202.38.79.95>
 - 端口号: 2121
 - 用户名、密码均为: sadb2011
 - 使用FTP工具上传
 - 提交内容按打包后按指定格式命名后再上传
- 实验提交情况
 - 1月3号会在网址<http://home.ustc.edu.cn/~linsh/addb.html>公布, 到时候大家可以去核实一下。

谢谢