

2022 年硕士研究生《分布式数据库》试题（共 100 分）

一、(20 分) 下面是某个公司人事数据库的两个全局关系

EMP={Eno, Ename, Title, Salary, Addr, Phone, Dno}; DEPT={Dno, Dname} 该公司共有 3 个部门, Dno 分别为 0, 1, 2。要求将 DEPT 关系和 EMP 关系的部分属性(Ename, Addr, Phone)保存在部门 0 的场地上, EMP 关系的部分属性(Title, Salary) 保存在所在部门场地上。根据上述要求,

1. 将全局模式进行分片, 写出分片定义和分片条件。
2. 指出各分片的类型, 并画出分片树。
3. 对查询 SELECT Ename, Salary, Dname FROM Emp, Dept WHERE Dno=2.进行全局优化, 画出优化后的全局查询树, 要求写出中间过程。
4. 进行分片优化, 画出优化后的分片查询树, 要求写出中间过程。

二、(20 分) 下面是当一个数据库系统出现故障时, 日志文件中的信息;

Crash

D ₀ ²	D ₁ ³	B ₂	D ₂ ¹	B ₃	C ₁	A ₃	B ₄	D ₂ ²	B ₅	K	D ₅ ¹	A ₅	C ₂	D ₄ ¹	A ₀	
-----------------------------	-----------------------------	----------------	-----------------------------	----------------	----------------	----------------	----------------	-----------------------------	----------------	---	-----------------------------	----------------	----------------	-----------------------------	----------------	--

根据上述 log 信息, 完成下面的处理:

1. 画出对应的事务并发执行图。
2. 说明检查点的作用和检查点时刻数据库需要完成的主要操作。
3. 确定出反做 (undo) 和重做 (redo) 事务集 (写出详细过程)。
4. 叙述 undo 和 redo 思想, 详细写出其基于日志的数据处理过程。

三、(30 分) 对 3 个关系 R,S 和 T 的分布式连接 $R \bowtie_{A=A} S \bowtie_{B=B} T$, 已知有如下的概要图:

Card(R) = 500

场地 S₁

Card(S) = 4000

场地 S₂

Card(T) = 200

场地 S₃

	A
Length	4
Val	500

	A	B
Length	4	4
Val	2000	1000

	B
Length	4
Val	200

假设通信代价系数 C₀= 0, C₁=1, DOM(R,A)⊆DOM(S,A), DOM(T,B)⊆DOM(S,B)。

按照 SDD-1 半连接优化算法, 逐步求出半连接优化集和最终执行场地。

四、(20 分) 设数据项 x, y 存放在 S₁ 场地, u, v 存放在 S₂ 场地。有分布式事务 T₁ 和 T₂, 判断下面的每个执行是否是局部可串行的, 是否是全局可串行的, 并分别说明事由。

1. 执行 1: 在 S₁ 场地 R₁(x)R₂(x)W₂(y)W₁(x), 在 S₂ 场地 R₁(u)W₁(u)R₂(v)W₂(u)。
2. 执行 2: 在 S₁ 场地 R₁(x)R₂(x)W₁(x)W₂(y), 在 S₂ 场地 W₂(u)R₁(u)R₂(v)W₁(u)。
3. 若在 S₁ 上有操作序列 R₁(x)W₁(x)R₁(y)W₁(y), 在 S₂ 场地上有操作序列

R₁(u)R₁(v)W₁(u), 假设 T₁ 的操作执行完成后将进行提交, 请按照 2PC 协议说明 T₁ 的提交处理过程, 并要求按照严格 2PL 协议, 对 T₁ 的操作处理加上显式的封锁操作和解锁操作。用 R₁₁(x)表示对 x 加读锁, W₁₁(x)表示对 x 加写锁, U₁₁(x)表示解锁。

五、(10 分) 理解分布式事务的 ACID 四个特性, 说明它同集中式数据库的异同。此外, 分布式事务的实现模型主要有哪些?