

## 2021 年硕士研究生《分布式数据库》试题（共 100 分）

一、(30 分)下面是某学校的学生关系数据库的全局模式：

某销售总公司（为场地 S0）下属有一个分公司（为场地 S1），该销售公司有如下关系模式：  
雇员信息：E（Eno（雇员编号），Ename（姓名），age（年龄），Title（级别），Dno（公司））

雇员销售信息：S（Sid（销售明细），Pno（商品编号），num（数量），date（日期），Eno）假设  
基于上面给出的关系模式有如下应用需求：（1）各分公司管理自己公司的雇员信息和销售信息；

（2）总公司管理 Title>5 的雇员信息和他们的销售信息。依据上述信息，

1. 进行数据分片设计，给出简单谓词集和小项谓词集，写出分片定义、分片条件。
2. 指出分片的类型，分别画出分片树。
3. 给出相应的分配设计，指出分配类型。
4. 要求查询 Title>8 的雇员信息和他们的销售信息（Eno（雇员编号），Ename（姓名），Pno（商品编号），num（数量），date（日期））。写出在全局模式上的 SQL 查询语句，并转换成相应的关系代数表达式，画出全局查询树。
5. 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。
6. 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 

$D_0^2$	$D_1^3$	$B_2$	$B_3$	$D_2^1$	$C_1$	$A_3$	K	$B_4$	$D_2^2$	$B_5$	$D_5^1$	$A_5$	$C_2$	$D_4^1$	$A_0$	
---------	---------	-------	-------	---------	-------	-------	---	-------	---------	-------	---------	-------	-------	---------	-------	--

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  $R \underset{B=B}{\bowtie} S \underset{C=C}{\bowtie} T$ ，已知有如下的概要图：

Card(R) = 300

场地 S<sub>1</sub>

	A	B
Length	20	10
Val	300	300

Card(S) = 4000

场地 S<sub>2</sub>

	B	C
Length	10	5
Val	1000	100

Card(T) = 50

场地 S<sub>3</sub>

	C
Length	5
Val	50

假设通信代价系数 C<sub>0</sub>= 0，C<sub>1</sub>=1，DOM(R,B)⊆DOM(S,B)，DOM(T,C)⊆DOM(S,C)。

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、(10 分) 某代售网站存在如下信息:航班信息表 Flight(Fno, Date, Price, Sold, Capacity, Co.), 客户信息表 Customer(Cid, Cname, Fno, CDate)。假设:场地 1 存储南航的航班信息 (Flight\_S) 和其客户信息(Customer\_S), 还有代收银行账户(B\_account);场地 2 存储国航的航班信息(Flight\_Z)和其客户信息 (Customer\_Z), 场地 3 存储用户账号(U\_account)。若用户 U 在该代售网站一次购买南航和国航混合航班的业务为一个分布式事务, 具体如下:步骤 1:客户 U 购买了南航航班 S 和国航航班 Z 各 1 张机票, 修改 Flight\_S 和 Flight\_Z 的售出数量(Sold);步骤 2:分别写用户 U 的南航购买信息 CS 和国航购买信息 CZ 到 Customer\_S 和 Customer\_Z 中;步骤 3:将用户 U 购买信息 CS 和 CZ 的花费 M 从用户账号(U\_account)转账到电商账户(B\_account)中。

若存在另一用户 V, 同时购买同 U 同样的航班, 也进行相同的操作, 请回答:

1. 几个场地上存在冲突操作?若存在冲突操作, 请写出冲突对;
2. 针对上述两个分布式事务, 请构建一全局历程 H, 同时写出: a) 局部可串行化、全局也可串行化的局部历程; b) 局部可串行化、全局不可串行化的局部历程。

五、（10 分）什么是数据库系统的“可靠性”和“可用性”？请举例说明二者的不同之处。