

2020 年计算机学院研究生《分布式数据库》试题（共 60 分）

1.(25 分)

某科研院由总部(A0)和两个下属分支机构(A1 和 A2)组成，总部和分支机构都从事科研业务。现对科研项目信息进行数据管理，涉及到的全局关系模式如下：

Branches(Bid, Bname, Adrs, Mgrno)，解释：机构基本信息表，其中：Bid 为机构编号，Bname 为机构名称，Adrs 是机构地址，Mgrno 是机构负责人员工编号；

Employees(Eid, Ename, Etitle, Bid)，解释：员工基本信息表，其中：Eid 为员工编号，Ename 为员工姓名，Etitle 为员工职务，Bid 是员工所属的机构编号；

Projects(Pid, Pname, Budget, DESC, Security, Pleader, Stime, Ftime)，解释：项目基本信息表，其中 Pid 为项目编号，Pname 为项目名称，Budget 为项目金额，DESC 为项目描述信息，Security 为项目是否涉密，Pleader 为项目负责人员工编号；

数据管理的基本需求是：(1)非涉密项目保存在项目负责人所在的分支机构场地，其中，金额超过 100 万元的项目信息在总部场地保存副本；(2)涉密项目信息只在总部场地保存；(3)总部和各分支结构保存各自的员工信息；(4)涉密项目的负责人信息在总部场地保存副本；(5)机构信息仅在总部场地保存。

请回答：

- 1) 按要求对表写出分片定义，画出分片树，并指明分片类型；
- 2) 按需求完成分配设计，指出分配类型；
- 3) 查询机构 A2 员工张龙负责的项目金额不超过 80 万的非涉密项目信息。要求查询结果包括张龙的姓名、项目名称、项目描述信息和项目金额。
- 4) 对 3)题的查询进行全局优化，要求写出中间过程；
- 5) 在全局优化的基础上进行片段优化，要求写出中间过程。

2.(10 分)

科研项目表存储在科技处场地，教师信息表存储在人事处场地，两个场地通过互联网连接。现从人事处场地查询计算机学院教师承担的项目信息。假设数据满足均匀分布，关系概要图如下：

科研项目表 Projects（单位：字节）						教师基本信息表 Faculties（单位：字节）				
Card(Projects)=30000 项, Length(Projects) = 400						Card(Faculties)=3000 人, Length(Employees) = 300				
	PID	FID	Pname	Sdate	...		FID	Fname	Institute	...
length	15	8	100	8	...	length	8	20	8	...
val	30000	2500	30000	25000	...	val	3000	2800	20	...
说明	项目编号	负责人教师编号	项目名称	立项时间	...	说明	教师编号	教师姓名	学院	...

计算：

- (1) 如果采用全连接优化算法，请给出查询执行计划，并计算通信代价。
- (2) 请给出优化的查询执行策略，并计算通信代价。

3.（8 分）

假设单主站点的情况下，有三个节点 A、B、C，其中 A 保存着 x 和 y 的主副本，B 和 C 保存着它们的副本，考虑下面事务：

T1

Read (x)

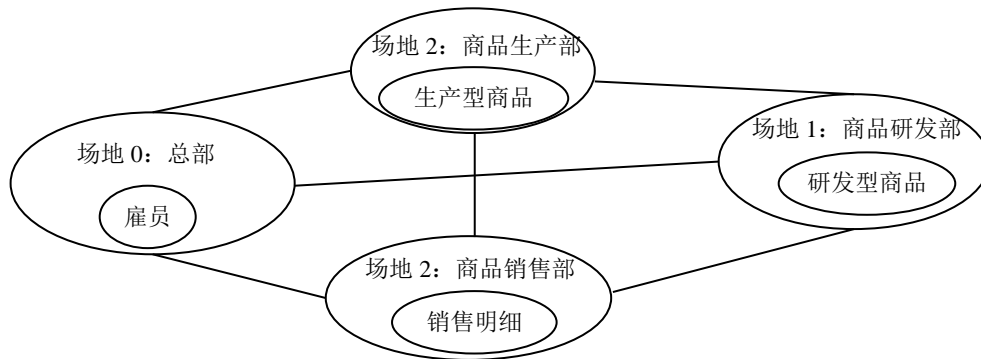
Write (y)

Commit

假设：T1 在 B 上提交，请回答：1) 结合同步复制协议的执行方式；2) 解释如何保证数据一致性。

4. (9 分)

某有限公司，根据业务需求设置了 4 个部门，分别为总部、商品研发部、商品生产部和商品销售部，分别在沈阳（场地 0）、北京（场地 1）、上海（场地 2）和广州（场地 3）四个城市，如下图所示。假设存在如下全局关系模式：雇员（雇员号，姓名，性别，部门），商品（商品号，商品名，品牌，商品类型，商品负责人），商品类型分为研发型和生产型，对应的负责人为研发负责人和生产负责人），销售明细（商品号，销售数量，销售人员）。



假设总部发布一个全局事务 T1:

Begin transaction T1

```
Select 商品类型, count (*)
from 商品
group by 商品类型
```

End

请回答:

- 1) 执行该全局事务时，请说明：该全局事务分为几个子事务；定义协调者和参与者，并说明它们完成的功能；
- 2) 若采用集中式两段提交协议提交，简述提交过程。
- 3) 在提交过程中，如果丢失了 commit 命令，如何处理？

5. (8 分)

接题 3，如果存在另两个全局事务 T2 和 T3:

Begin transaction T2

```
Insert into 商品 values ("010", "R010", " ", "研发型", "王一")
```

End

Begin transaction T3

```
Insert into 商品 values ("110", "B110", " ", "生产型", "李二")
```

End

假设：T1、T2、T3 并发执行，

请回答:

- 1) 指出存在冲突操作的场地，并写出存在的冲突对；
- 2) 构建一个局部可串行化，而全局不可串行化的调度历程，并说明为什么？