1.试述查询优化在关系数据库系统中的重要性和可能性

答：重要性:关系系统的查询优化既是 RDBMS 实现的关键技术,又是关系系统的优点所在，大学搬题器它减轻了用户选择存取路径的负担,用户只要提出“千什么”,不必指出“怎么干”查询优化的优点不仅在于用户不必考虑如何最好地表达查询以获得较好的效率,而且在于系统可以比用户程序的“优化”做得更好。

可能性:

(1)优化器可以从数据字典中获取许多统计信息,例如各个关系中的元组数,关系中每个属性值的分布情况,这些属性上是否有索引、是什么索引(B+树索引、Hash 索引、唯一索引还是组合索引),等等。优化器可以根据这些信息选择有效的执行计划,而用户程序则难以获得这些信息。

(2)如果数据库的物理统计信息改变了,系统可以自动对查询进行重新优化以选择相适应的执行计划。在非关系系统中必须重写程序,而重写程序在实际应用中往往是不太

可能的。

(3)优化器可以考虑数十甚至数百种不同的执行计划,从中选出较优的一个,而程序员-般只能考虑有限的几种可能性。

(4)优化器中包括了很多复杂的优化技术,这些优化技术往往只有最好的程序员才能掌握。系统的自动优化相当于使得所有人都拥有这些优化技术。

2.假设关系R(A,B)和S(B,C,D)情况如下:R有20 000 个元组,S有1200个元组,个块能装 40个R的元组,能装 30个S 的元组估算下列操作需要多少次磁盘块读。

(1)DR上没有索引,select \* from R;

(2)R中A为主码,其上有3层B+树索引,select \* from R where A=10;

(3)套循环连接R S;

(4)排序合并连接 R S,区分R与 S 在B属性上有序和无序两种情况

答：

(1)需要对R进行全表扫描块数=20000/40=500

(2)对R进行索引扫描块数=3+1=4;其中3块B+树索引块1块数据块。

(3)R本身20000/40=500个块S本身1200/30=40个块以S为外表假设内存分配的块

数为k,嵌套循环连接需要的块数为40+(40/(k-1))\*500

(3)如果R和S都在B属性上排好序块数500+40=540;如果都没有排序则还要加上排序代价,结果为540+2\* 500 \* (log2500+1)+2\*40\*(log240+1)。

3.对学生-课程数据库,查询信息系学生选修了的所有课程名称。

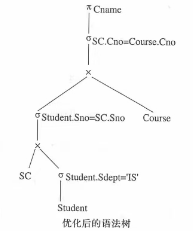
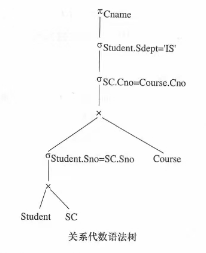
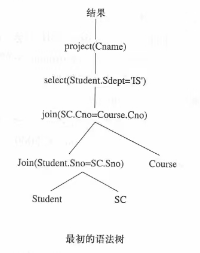
SELECT Cname

FROM Student,Course,SC

WHERE Student.Sno=SC.Sno AND SC.Cno= Course.Cno AND Student.Sdept='IS'

此查询要求信息系学生选修了所有课程名称。试画出用关系代数表示的语法树,并用关系代数表达式优化算法对原始的语法树进行优化处理,画出优化后的标准语法树。

答：



4.对于下面的数据库模式:

Teacher(Tno,Tname,Tage,Tsex);

Department(Dno,Dname,Tno);

Work(Tno, Dno, Year,Salary)

假设 Teacher的 Tno属性、Department 的 Dno属性以及 Work 的 Year 属性上有 B+树索引,说明通常情况下列操作的处理方法。

(1)select \* from teacher where tsex='女'

(2)select \* from department where dno < 301

(3)select \* from work where year <> 2000

(4)select \* from work where year > 2000 and salary < 5000

(5)select \* from work where year < 2000 or salary < 5000

答：

(1)对 teacher 进行全表扫描,查看元组是否满足性别为女。

(2)如果满足 dno< 301 的元组数目较少,可以通过索引找到dno=301 的索引项,然后顺着 B+树的顺序集得到 dno< 301 的索引项,通过这些指针找到 department 中的元组;如果

dno<301的元组数目较多,可以采用对 department 的全表扫描方式处理。

(3)对 work 进行全表扫描,查看元组是否满足 year <> 2000。

(4)通过year 的索引找到满足 year > 2000的元组检查元组是否满足 salary < 5000

(5)对 work 进行全表扫描查看元组是否满足 year < 2000或 salary <5000。

5.对于题 4 中的数据库模式,存在如下的查询:

SELECT tnam

FROM teacher,department, work

WHERE teacher.tno= work.tno AND department.dno = work.dno AND

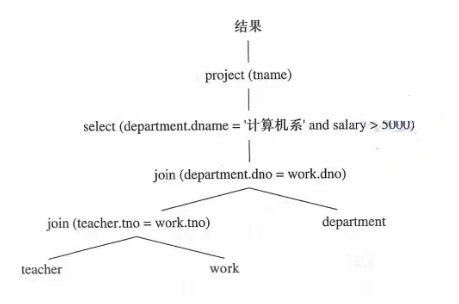
department.dname='计算机系AND salary >5000

画出语法树以及用关系代数表示的语法树,并对关系代数语法树进行优化,画出优化后

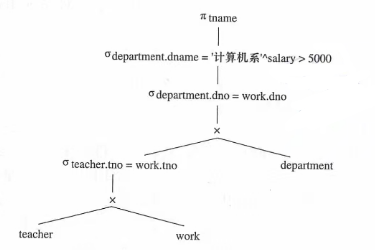
的语法树。

答：

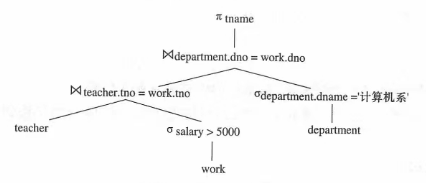
语法树：



关系代数语法树：



优化后的语法树：



6.试述关系数据库管理系统查询优化的一般准则。

答：下面的优化策略一般能提高查询效率：

(1)选择运算应尽可能先做。

(2)把投影运算和选择运算同时进行。

(3)把投影同其前或其后的双目运算结合起来执行。

(4)把某些选择同在它前面要执行的笛卡儿积结合起来成为一个连接运算。

(5)找出公共子表达式。

(6)选取合适的连接算法。

7.试述关系数据库管理系统查询优化的一般步骤。

答:

各个关系系统的优化方法不尽相同,大致的步骤可以归纳如下:

(1)把查询转换成某种内部表示,通常用的内部表示是语法树。

(2)把语法树转换成标准(优化)形式,即利用优化算法把原始的语法树转换成优化的

形式。

(3)选择低层的存取路径。

(4)生成查询计划,选择所需代价最小的计划加以执行。