

数据库系统2-7：查询实例相关 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/134/2021_2022__E6_95_B0_E6_8D_AE_E5_BA_93_E7_c97_134643.htm 计算机等级考试训练

软件《百宝箱》查询有多种实现方法，例如查“王平”所选修的课程编号及成绩，系统可用多种等价的关系代数表达式完成这一操作，如 $T1 = \pi_{CNO, GRADE}(\sigma_{S.SNO=S.C.SNO \wedge S.SNAME = \text{王平}}(S \bowtie SXS_C))$

$T2 = \pi_{CNO, GRADE}(\sigma_{SNAME = \text{王平}}(S \bowtie SXS_C))$ $T3 = \pi_{CNO, GRADE}(\sigma_{SNAME = \text{王平}}(S)) \bowtie SXS_C$ 在这些操作中，它们的结果是一样的，但执行过程相差很大，且系统的开销也大相径庭。对于T1而言，当计算SXS_C时，需要把S和S-C的全部元组连接起来。假设每个存储块能保存10元组，S的物理文件在存储器中需B1个存储块，S_C的物理文件需B2块。内存中提供的运算缓冲区最多能装n块，而B1、B2均大于n。因而对乘积较好的执行办法是：将S文件分成若干个n-1块，首先将第一个n-1块装入内存，并逐次装入S_C文件的一个块，使之与S已装入的n-1块进行乘积运算；当S_C文件的每块都装入一遍后，再往内存装入S文件的下一个n-1块，并同样从第一块开始逐次地装入S_C的每一块，重复执行上述连接运算，这样直到计算完乘积的全部元组为止，其读块数目为： $B1 \lceil B1 / (n-1) \rceil * B2$

设 $B1=B2=1500$ ， $n=80$ ，则所需的读块的总数目为1500

$\lceil 1500/79 \rceil * 1500 = 30000$ 。假设一秒钟能读20块时，大约需要25分钟时间。同样假设每个存储块能保存联接后的10个元组和一秒钟能写入20块，则联接后一共有 $1500 * 1500 = 2250000$ 块，将联接后的中间结果写入存储器需要1875分钟，然后再将中

间结果读出来进行选择 and 投影，也需1875分钟。与读块和写块相比，联接、选择、投影等运算时间均可忽略不计。完成T1表达式的运算的时间大约需要3775分钟，即62小时。对T3而言，先对S文件作SNAME= " 王平 " 的选择操作，读块数目为B1；然后把结果与S_C作连接、投影运算，读块数目为B2，所以总的读块数目为B1 B2=3000，由于满足条件的元组很少(大约50个元组)，不用保存中间结果文件，因而完成T3表达式共需约2.5分钟(一秒钟仍读20块)，仅等于前者的一千五百分之一。当文件的存贮块数更多且存在关于SNAME的倒排索引时，两者的时间差别将更为显著。对于一个运算表达式，能否找出一个与之等价且操作时间更少的表达式呢？这正是查询优化所要研究的问题。

100Test 下载频道开通，各类考试题目直接下载。详细请访问
www.100test.com