

## 数据库系统课程教案

## 第 20 单元

学时：2

教材内容	第 9 章 关系查询处理和查询优化
基本知识 点	关系系统的定义、关系系统的分类；查询优化的概念、查询优化的一般准则、关系代数等价变换规则及优化算法、画出一个查询的语法树以及优化后的标准语法树
教学重点	查询优化的概念、查询优化的一般准则、关系代数等价变换规则及优化算法、画出一个查询的语法树以及优化后的标准语法树
教学难点	画出一个查询的语法树以及优化后的标准语法树
要求掌握 内容	最小关系系统、关系完备系统、全关系系统和关系系统查询优化的概念、画出一个查询的语法树以及优化后的标准语法树
教学思路， 采用的教学 方法和 辅助手段， 板书设计， 重点如何 突出，难点 如何解决， 师生互动 等	<p><b>教学思路：</b></p> <p><b>一、复习旧课，巩固上次授课主要内容</b></p> <p>1、简述数据库设计的六个步骤。</p> <p>2、简述 E-R 图向关系模型转换的原则。</p> <p><b>二、导入新课，明确本次授课的目的与要求</b></p> <p>1、理解关系系统的定义。</p> <p>2、了解关系系统的分类。</p> <p>3、掌握关系系统查询优化的步骤。</p> <p>4、掌握关系系统查询优化的一般准则。</p> <p><b>三、讲解本次授课的具体内容</b></p> <p><b>教学方法：</b>整合教学内容，强化基础训练；努力营造生动活泼的课堂气氛，搭建师生间良好的沟通渠道；采用多媒体教学与传统的板书设计相结合的方式，教学手段灵活多变。</p> <p><b>辅助手段：</b>通过 PPT 幻灯片演示结合板书设计和举例加以阐述。</p>
本章思考题 和作业	P290 第 2、3、4、5 题
主要教材 参考资料	<p>1. 《数据库系统概论》，萨师煊，王珊，高等教育出版社，2014.9</p> <p>2. 《数据库系统概论学习指导与习题解答》，王珊，张俊，高等教育出版社，2015.7</p>
备 注	

## 本次授课具体内容

## 第 9 章 查询处理及其查询优化

### 9.1 关系数据库系统的查询处理关系数据库管理系统查询处理阶段：

#### 1. 查询分析

查询分析的任务：对查询语句进行扫描、词法分析和语法分析

词法分析：从查询语句中识别出正确的语言符号

语法分析：进行语法检查

#### 2. 查询检查

查询检查的任务

合法权检查

视图转换

安全性检查

完整性初步检查

根据数据字典中有关的模式定义检查语句中的数据库对象，如关系名、属性名是否存在和有效

如果是对视图的操作，则要用视图消解方法把对视图的操作转换成对基本表的操作

#### 3. 查询优化

查询优化：选择一个高效执行的查询处理策略

查询优化分类

代数优化/逻辑优化：指关系代数表达式的优化

物理优化：指存取路径和底层操作算法的选择

查询优化的选择依据

基于规则(rule based)

基于代价(cost based)

基于语义(semantic based)

#### 4. 查询执行

依据优化器得到的执行策略生成查询执行计划

代码生成器(code generator)生成执行查询计划的代码

两种执行方法

自顶向下

自底向上

9.2 关系系统的查询优化(一)查询优化的概述：关系查询优化是影响 RDBMS 性能的关键因素，由于关系表达式的语义级别很高，使关系系统可以从关系表达式中分析查询语义，提供了执行查询优化的可能性。

(二)查询优化的优点：用户不必考虑如何最好地表达查询就可以获得较好的效率，系统优化比用户程序优化做得更好，这是因为：

1、优化器可以从数据字典中获取许多统计信息，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、优化器可以考虑数百种不同的执行计划，而程序员一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。(三)查询优化的目标：选择有效策略，求得给定关系表达式的值。

(四)查询优化的步骤

1、将查询转换成某种内部表示，通常是语法树。 2、根据一定的等价变换规则把语法树转换成

标准(优化)形式。

3、选择低层的操作算法。对于语法树中的每一个操作，计算各种执行算法的执行代价，选择代价小的执行算法。

4、生成查询计划(查询执行方案)：由一系列内部操作组成的。

(五)代价模型

1、集中式数据库：单用户系统的总代价 = I/O 代价 + CPU 代价

多用户系统的总代价 = I/O 代价 + CPU 代价 + 内存代价

2、分布式数据库：总代价 = I/O 代价 + CPU 代价[+ 内存代价] + 通信代价

(六)查询优化的必要性：一个实例：求选修课程'2'的学生姓名，用 SQL 语言表达为：

SELECT Sname FROM Student, SC WHERE Student.Sno=SC.Sno AND SC.Cno='2' ;

假设学生~课程数据库有 1000 个 Student 记录、10000 个 SC 记录，其中选修 2 号课程的选课记录为 50 条。对于上述 SQL 查询，DBMS 可以转化为如下几种等价的关系代数表达式，由于执行策略不同，查询的时间相差很大。

1、执行策略 1： $\Pi_{Sname}(\sigma_{Student.Sno=SC.Sno \wedge SC.Cno='2'}(Student \times SC))$

(1)计算广义笛卡尔积：把 Student 和 SC 的每个元组连接起来。

方法是：在内存中尽可能多地装入 Student 表的若干块元组，留出一块存放 SC 表的一块元组。然后把 SC 中的每个元组和 Student 中每个元组连接，连接后的元组装满一块后就写到中间文件上，再从 SC 中读入一块和内存中的 Student 元组连接，直到 SC 表处理完。再次读入若干块 Student 元组，读入一块 SC 元组，重复上述处理过程至 Student 表处理完毕。假设一个内存块能装 10 个 Student 元组或 100 个 SC 元组，内存中一次可以存放 5 块 Student 元组和 1 块 SC 元组及若干块连接结果元组，则

读取总块数 = 读 Student 表块数 + 读 SC 表遍数  $\times$  每遍块数  
 $= 1000/10 + (1000/(10 \times 5)) \times (10000/100) = 2100$

若每秒读写 20 块，则读块总计时间 =  $2100/20 = 105$  秒。连接后的元组数为  $1000 \times 10000 = 10^7$ ，设每块能装 10 个这样的元组，则写出这些块总计时间 =  $10^7/10/20 = 50000$  秒。(2)作选择操作：忽略内存处理时间，这一步共计时间同写中间文件 50000 秒。(3)作投影操作：时间可以忽略。

(4)结论：查询总时间 =  $(105 + 50000 + 50000)$  秒 = 100105 秒 = 27.8 小时。执行策略 2： $\Pi_{Sname}(\sigma_{SC.Cno='2'}(Student \bowtie SC))$

(1)作自然连接：为了执行自然连接，读取 Student 和 SC 表的策略不变，读取总块数 2100，共计时间 105 秒。但自然连接的结果为 10000，比策略 1 大大减少了，写出这些元组时间为  $10000/10/20 = 50$  秒。(2)作选择操作：读取中间文件块，作选择运算共计时间 50 秒。(3)作投影操作：时间可以忽略。

(4)结论：查询总时间 =  $(105 + 50 + 50)$  秒 = 205 秒 = 3.4 分钟。

### 9.3 代数优化

(一)代数优化策略：通过对关系代数表达式的等价变换来提高查询效率。

关系代数表达式的等价：指用相同的关系代替两个表达式中相应的关系所得到的结果是相同的。常用的等价变换规则。

(二)查询树的启发式优化

典型的启发式规则：

(1)选择运算应尽可能先做

在优化策略中这是最重要、最基本的一条。

(2)把投影运算和选择运算同时进行

如有若干投影和选择运算，并且它们都对同一个关系操作，则可以在扫描此关系的同时完成所有的这些运算以避免重复扫描关系。

(3)把投影同其前或其后的双目运算结合起来，没有必要为了去掉某些字段而扫描一遍关系。

(4) 把某些选择同在它前面要执行的笛卡尔积结合起来成为一个连接运算，连接特别是等值连接运算要比同样关系上的笛卡尔积省很多时间。

(5) 找出公共子表达式

如果这种重复出现的子表达式的结果不是很大的关系，并且从外存中读入这个关系比计算该子表达式的时间少得多，则先计算一次公共子表达式并把结果写入中间文件是合算的。

当查询的是视图时，定义视图的表达式就是公共子表达式的情况。

#### 9.4 物理优化

对于一个查询语句有许多存取方案，它们的执行效率不同，仅仅进行代数优化是不够的，物理优化就是要选择高效合理的操作算法或存取路径，求得优化的查询计划。

物理优化方法：

##### 1、基于规则的启发式优化

启发式规则是指那些在大多数情况下都适用，但在不是每种情况下都是适用的规则。

##### 2、基于代价估算的优化

优化器估算不同执行策略的代价，并选出具有最小代价的执行计划。

##### 3、两者结合的优化方法：

常常先使用启发式规则，选取若干较优的候选方案，减少代价估算的工作量，然后分别计算这些候选方案的执行代价，较快地选出最终的优化方案。

### 本次授课小结

本次授课讲述了关系系统的定义、关系系统的分类、关系系统查询优化的步骤、关系系统查询优化的一般准则。

学生课后复习时应着重于其中的第 1、3、4 点内容，为进一步学习后续章节打好基础。

实验

无