第三篇　系统篇

- 讨论数据库管理系统中查询处理和事务管理的基本概念和基础知识
  - 关系查询处理和查询优化
  - 数据库恢复
  - 并发控制

数据库系统概论
An Introduction to Database System

第九章 关系系统及其查询优化

中国人民大学信息学院 陈红

9.1 关系数据库系统的查询处理

9.2 关系数据库系统的查询优化

9.3 代数优化

9.4 物理优化

9.5 查询执行

9.6 小 结

9.3 代 数 优 化

- 9.3.1 关系代数表达式等价变换规则
- 9.3.2 查询树的启发式优化

9.3.1 关系代数表达式等价变换规则

- 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
- 两个关系表达式 $E1$ 和 $E2$ 是等价的，可记为 $E1 \equiv E2$

关系代数表达式等价变换规则（续）

❖ 常用的等价变换规则：

1. 连接、笛卡尔积交换律

设 $E1$ 和 $E2$ 是关系代数表达式， F 是连接运算的条件，则有

$$E_1 \times E_2 \equiv E_2 \times E_1$$

$$E_1 \bowtie E_2 \equiv E_2 \bowtie E_1$$

$$E_1 \underset{F}{\bowtie} E_2 \equiv E_2 \underset{F}{\bowtie} E_1$$

2. 连接、笛卡尔积的结合律

设 E_1， E_2， E_3 是关系代数表达式， F_1 和 F_2 是连接运算的条件，则有

$$(E_1 \times E_2) \times E_3 \equiv E_1 \times (E_2 \times E_3)$$

$$(E_1 \underset{F_1}{\bowtie} E_2) \underset{F_2}{\bowtie} E_3 \equiv E_1 \underset{F_2}{\bowtie} (E_2 \underset{F_2}{\bowtie} E_3)$$

$$(E_1 \quad E_2) \quad E_3 \equiv E_1 \quad (E_2 \quad E_3)$$

关系代数表达式等价变换规则（续）

3. 投影的串接定律

$$\pi_{A_1,A_2,\cdots A_n}(\pi_{B_1,B_2,\cdots B_m}(E)) \equiv \pi_{A_1,A_2,\cdots A_n}(E)$$

这里，E 是关系代数表达式，A_i(i=1，2，…，n)，B_j(j=1，2，…，m) 是属性名且 {A_1，A_2，…，A_n} 构成 {B_1，B_2，…，B_m} 的子集。

4. 选择的串接定律

$$\sigma_{F_1}(\sigma_{F_2}(E)) \equiv \sigma_{F_1 \wedge F_2}(E)$$

这里，E 是关系代数表达式，F_1、F_2 是选择条件。

选择的串接律说明选择条件可以合并。这样一次就可检查全部条件。

关系代数表达式等价变换规则（续）

5. 选择与投影操作的交换律

$$\sigma_F(\pi_{A_1,A_2,\cdots A_n}(E)) \equiv \pi_{A_1,A_2,\cdots A_n}(\sigma_F(E))$$

选择条件 F 只涉及属性 A_1，…， A_n。

若 F 中有不属于 A_1，…， A_n 的属性 B_1，…， B_m 则有更一般的规则：

$$\pi_{A_1,A_2,\cdots A_n}(\sigma_F(E)) \equiv \pi_{A_1,A_2,\cdots A_n}(\sigma_F(\pi_{A_1,A_2,\cdots A_n,B_1,B_2,\cdots B_m}(E)))$$

关系代数表达式等价变换规则（续）

6. 选择与笛卡尔积的交换律

如果 F 中涉及的属性都是 E_1 中的属性，则

$$\sigma_F(E_1 \times E_2) \equiv \sigma_F(E_1) \times E_2$$

如果 $F=F_1 \wedge F_2$，并且 F_1 只涉及 E_1 中的属性，F_2 只涉及 E_2 中的属性，则由上面的等价变换规则 1，4，6 可推出：

$$\sigma_F(E_1 \times E_2) \equiv \sigma_{F_1}(E_1) \times \sigma_{F_2}(E_2)$$

若 F_1 只涉及 E_1 中的属性，F_2 涉及 E_1 和 E_2 两者的属性，则仍有

$$\sigma_F(E_1 \times E_2) \equiv \sigma_{F_2}(\sigma_{F_1}(E_1) \times E_2)$$

它使部分选择在笛卡尔积前先做。

关系代数表达式等价变换规则（续）

7. 选择与并的分配律

设 $E=E_1 \cup E_2$， E_1， E_2 有相同的属性名，则

$\sigma_F(E_1 \cup E_2) \equiv \sigma_F(E_1) \cup \sigma_F(E_2)$

8. 选择与差运算的分配律

若 E_1 与 E_2 有相同的属性名，则

$\sigma_F(E_1 - E_2) \equiv \sigma_F(E_1) - \sigma_F(E_2)$

9. 选择对自然连接的分配律

$\sigma_F(E_1 \bowtie E_2) \equiv \sigma_F(E_1) \bowtie \sigma_F(E_2)$

F 只涉及 E_1 与 E_2 的公共属性

关系代数表达式等价变换规则（续）

10. 投影与笛卡尔积的分配律

设 E_1 和 E_2 是两个关系表达式，A_1，…，A_n 是 E_1 的属性，B_1，…，B_m 是 E_2 的属性，则

$$\pi_{A_1,A_2,\cdots A_n,B_1,B_2,\cdots B_m}(E_1 \times E_2) \equiv \pi_{A_1,A_2,\cdots A_n}(E_1) \times \pi_{B_1,B_2,\cdots B_m}(E_2)$$

11. 投影与并的分配律

设 E_1 和 E_2 有相同的属性名，则

$$\pi_{A_1,A_2,\cdots A_n}(E_1 \cup E_2) \equiv \pi_{A_1,A_2,\cdots A_n}(E_1) \cup \pi_{A_1,A_2,\cdots A_n}(E_2)$$

小结

1-2:　连接、笛卡尔积的交换律、结合律

3：　合并或分解投影运算

4：　合并或分解选择运算

5-9：　选择运算与其他运算交换

5,10,11：　投影运算与其他运算交换

9.3 代 数 优 化

- 9.3.1　关系代数表达式等价变换规则

- 9.3.2　查询树的启发式优化

9.3.2 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

查询树的启发式优化（续）

- 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。

查询树的启发式优化（续）

- 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

查询树的启发式优化（续）

- 遵循这些启发式规则，应用 9.3.1 的等价变换公式来优化关系表达式的算法。

  算法：关系表达式的优化

  输入：一个关系表达式的查询树

  规则 4：　合并或分解选择运算
  规则 5-9：　选择运算与其他运算交换

  (1) 利用等价变换规则 4 把形如 $\sigma_{F1 \wedge F2 \wedge \ldots \wedge Fn}(E)$ 变换为 $\sigma_{F1}(\sigma_{F2}(\ldots(\sigma_{Fn}(E))\ldots))$。

  (2) 对每一个选择，利用等价变换规则 4～9 尽可能把它移到树的叶端。

查询树的启发式优化（续）

(3) 对每一个投影利用等价变换规则 3， 5， 10， 11 中的一般形式尽可能把它移向树的叶端。

- 注意：

规则 3： 合并或分解投影运算
规则 4： 合并或分解选择运算
规则 5： 投影运算与选择运算交换

一些投影出现

一个有可能被

规则 3： 合并或分解投影运算
规则 5,10,11： 投影运算与其他运算交换

的串接合并成单

影。使多个选择

完成

查询树的启发式优化（续）

(5) 把上述得到的语法树的内节点分组。每一双目运算 ($\bowtie$， ，∪， -) 和它所有的直接祖先为一组 (这些直接祖先是 (σ， π 运算) 。

- 如果其后代直到叶子全是单目运算，则也将它们并入该组
- 但当双目运算是笛卡尔积 (×)，而且后面不是与它组成等值连接的选择时，则不能把选择与这个双目运算组成同一组

查询树的启发式优化（续）

［例 4］ 下面给出［例 3］中 SQL 语句的代数优化示例。

(1) 把 SQL 语 ，如下图所示

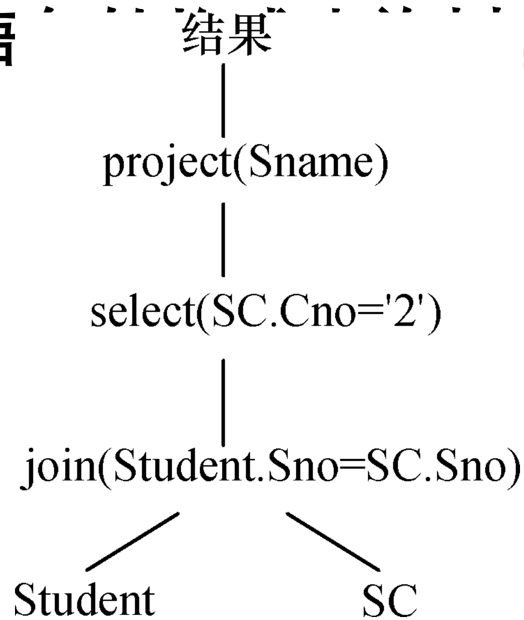
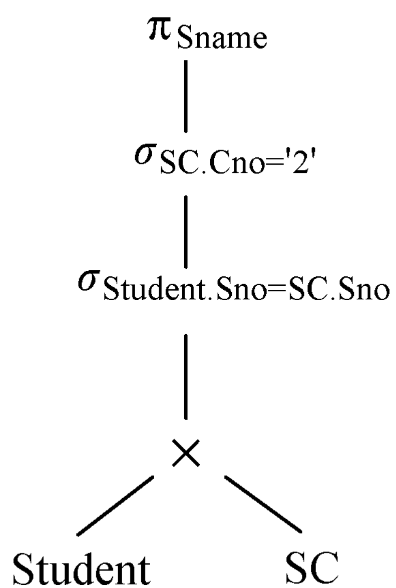


图 9.4 查询树

查询树的启发式优化（续）

为了使用关系代数表达式的优化法，假设内部表示是关系代数语法树，则上面的查询树如下图所示。



关系代数语法树

查询树的启发式优化（续）

(2) 对查询树进行优化

利用规则 4、 6 把选择 $\sigma_{SC.Cno}$='2' 移到叶端，图 9.4 查询树便转换

成图 9.5 优化的查询树。这就是 9.2.2 节中 Q3 的查询树表示

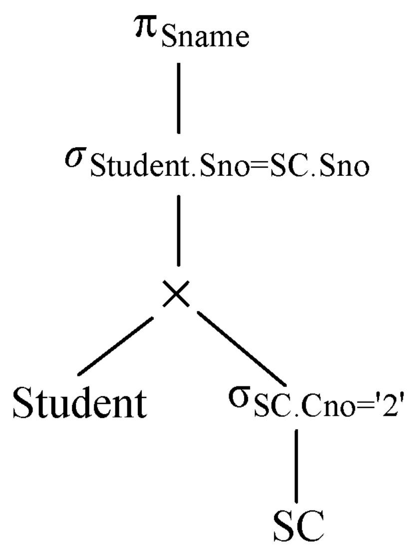


图 9.5 优化后的查询树