

数据库第九章

查询处理

任务

把用户提交给关系数据库管理系统的查询语句转换为高效的查询执行计划

步骤/阶段

查询分析

词法分析

从查询语句中识别出正确的语言符合

语法分析

进行语法检查

查询检查

概念

选择一个高校执行的查询处理策略

概念

关系代数表达式的优化

按照一定的规则 通过对关系代数表达式进行等价变换 改变代数表达式中操作的次序和组合 使查询执行更高效

方法

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

概念

用相同的关系代替两个表达式中的相应地关系 所得到的结果是相同的

查询树的启发式优化

概念

通过存取路径和底层操作算法的选择进行的优化

概念

启发式优化是指 在大多数情况下都适用 但不一定是最好的规则

分类

代数优化

物理优化

基于规则的启发式优化

规则

选择操作的启发式规则

概括

选择操作只涉及一个关系 一般采用全表扫描或者基于索引的算法

当选择率比较低时 索引扫描

当选择率比较高/待查找的元组分布均匀 全表扫描

详细

对于小关系 使用全表顺序扫描 即使选择列上有索引

对于主码=值的查询 可以选择主码索引

对于非主属性=值的查询 并且选择列上有索引 如果比例较小（低于10%）可以选择索引扫描 否则使用全表扫描

对于非等值查询或者范围查询 且选择列上有索引 如果选择率低于10% 可以使用索引扫描 否则使用全表顺序扫描

对于用AND连接的合取选择条件 如果涉及属性的组合索引 则采用组合索引扫描 如果某些属性上有一般的索引 则使用索引扫描 否则全表顺序扫描

对于用OR连接的析取选择条件 一般使用全表顺序扫描

连接操作的启发式规则

排序—合并算法

如果两个表都按照连接属性排序

索引连接算法

如果一个表在连接属性上有索引

hash join算法

如果上面两个规则都不适用 且其中一个表较小

嵌套循环算法

处理等值连接的算法

最后可以选用嵌套循环算法 并选择其中较小的表作为外表

最简单最普通的连接方法 可以处理包括非等值连接在内的各种连接操作

基于代价估算的优化

概念

选出具有最小代价的执行计划

代价估算

全表扫描算法 若基本表大小为B块 全表扫描算法的代价cost=B

索引扫描算法 若选择条件是码=值 那么平均搜索代价cost=B/2

嵌套循环连接算法

排序—合并连接算法

两者结合的优化方法

基于语义的优化

步骤

把查询转换成某种内部表示 通常用的内部表示是语法树

把语法树转换成标准形式 即利用优化算法 把原始的语法树转换成优化的形式

选择低层的存取路径

生成查询计划 选择代价最小的

优化代价

在集中式数据库中

单用户

总代价=I/O代价+CPU代价

多用户

查询执行开销=磁盘存取块数（I/O代价）+处理机时间（CPU代价）+查询的内存开销

在分布式数据库中

总代价=I/O代价+CPU代价+内存代价+通信代价

备注

计算查询代价时 一般用查询处理读写块数作为衡量单位

总目标

选择有效的策略 求得给定关系的表达式的值 使得查询代价较小

查询执行

由代码生成器生成执行这个查询计划的代码 然后加以执行 回送查询结果



扫码听课，视频讲解更清晰