# 实验报告

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 《数据库系统》——实验报告 | | | | | |
| 姓名： | 陈卓文 | 班级： | 19级2班 | 学号： | 201936380215 |
| 题目： | 实验五 查询优化 | | | 日期： | 2021.6.7 |
| 实验环境：  Ubuntu 18.04  Mysql 8.0.25 | | | | | |
| 实验内容与完成情况：  1 建立关系表，并建立索引；导入数据。  a) 建立索引。  1) 在四张表的主码上建立聚簇索引；    2) 自选在一些属性上建立非聚簇索引；    3) 建立索引时，请思考索引的填充因子是什么意思，有什么作用；  请思考索引的数据结构有哪些？可否指定索引的数据结构？如果不可以，请说明为什么？如果可以，请说明你的操作过程或SQL操作。  答：填充因子指定在创建索引后对数据页的填充比例，值越小，则数据页上空闲空间越大，减少在索引增长过程中对数据页拆分的需要。能进行高效索引的数据结构有：哈希表、二叉树、B树、B+树等。不能指定，索引结构是默认实现的。  b) 导入空数据。  请同学们在db\_SPJ中自行增加一些有空值的元组。    2 简单查询  a) 在单表上进行查询，查看某个或多个特定属性.    b) 使用排序操作Order By    c) 使用分组操作Group by、having；并作相关的集函数查询；例如sum，count，avg，max，min等    d) 取消重复distinct    e) 通配符的使用。    对于以上查询，请大家仔细看各个查询的查询计划图，回答以下问题并举出实例：   1. 空值对结果的影响（例如在计算sum, avg, min集函数时系统如何处理空值属性）？   答：空值不影响结果，处理时直接被忽略了。   1. 注意察看查询执行计划：如果察看的属性列上有索引和没有索引，系统是读取数据的方式有何不同？   答：当没有索引时，读取数据type是all，即遍历全表；有索引type为index，仅遍历索引。   1. 是不是在一个表上有索引，所有的查询都会使用该索引去读取数据？为什么？   答：不一定，当查询涉及的属性与索引属性相关时，才能使用对应的索引来加速查找。   1. 为什么有时候即使使用了Order by操作，但查询计划里并没有相应的操作？对于distinct也有类似的情况。   答：因为order by的属性有对应索引，已经有序了，不必再进行排序，去重同理。  3 复杂的查询  a) 连接查询。  1) 观察查询计划图，Join操作系统有哪些不同的实现join的方式？    2) 系统是否区别等值连接和自然连接？如果不区分，为什么？如果区分，请说明系统是如何区分的？  答：不区分，自然连接就是特殊的等值连接。  b) 嵌套查询  1) 对同一个查询，写出非嵌套查询和嵌套查询两种形式；观察他们的执行速度差异；观察他们的执行计划有何不同，并解释为什么。    答：嵌套查询需要提前进行一次查询得到子查询的结果。  2) 对同一个查询，写出相关嵌套查询和非相关嵌套查询的形式，执行计划和执行效率对比    答：两个执行计划是一样的，两者可以互相转化，具体执行过程被优化过。  3) 通过查询验证IN，Exists，ANY，ALL，Some之间的等价关系  答：由上条可知 in 和 exists 等价    IN    ALL    SOME  c) 复合条件查询  1) Where条件中除了有连接条件外，还有一些选择条件。    2) 注意察看它们的执行计划，执行这些操作的顺序，是否是安装语句中条件的顺序来执行，是先做选择还是先做连接？  答：并非按照语句中的条件的顺序来执行，而是先做选择，再做连接。  思考和要求：   1. 对于复杂的查询，尽量使用不同的形式表示出来。对比其结果，是否一样？查询计划是否一样？   答：在3.b.2中几条语句的结果相同，但是查询计划不一样   1. 对同样的查询语句，有无索引的情况下，他们的查询计划有何不一样？   答：在存在索引的情况下，查询的type是index，否则为all，两者存在较大的效率差距  4 引入视图  实验步骤3中的查询，先建立相应的视图。再在视图上进行查询。观察他们结果是否一样。它们的执行计划是否一样。分别给出例子。    答：可以看出使用视图后的执行计划仍然一样，说明每次查询时，数据库会重建数据。 | | | | | |