# 实验7-索引

课程名称：数据库系统

实验学时：16 学时

综合性、设计性实验：□🗹是 □否

面向专业和班级： 软件工程

## 任务编制人：曾兵

## 一、实验目的

掌握索引的创建和使用, 体会索引的高效.

## 二、实验内容

1. 运行脚本largeRelationsInsertFile.sql (见教学在线) 创建一个名为university数据库及其表格和数据.
2. 任选以上步骤中创建好的表格, 设计查询语句, 对比有索引和没有索引在查询效率上的区别. 为了使区别变得明显, 你可能需要添加更多的数据.
3. 填写如下表格

|  |  |
| --- | --- |
| 内存大小 |  |
| 数据量大小 |  |
| 外存类型(磁盘or 固态硬盘) |  |
| 数据库版本 |  |
| 操作系统版本 |  |
| 编程语言 |  |
| 重复实验的次数 |  |
| 有索引的时间/无索引的时间 |  |

## 三、实验环境

本实验推荐如下环境.

1. 操作系统: Windows 2000 及其以上
2. 数据库: SQL service 2008 及其以上

本实验也可在其它任意硬件机器, 任意关系数据库上进行.

## 四、重要提示

度量有无索引在查询效率上的差异, 并不是一件简单的事情。假设无索引的时间开销是t1, 有索引的时间开销是t2. 一次查询中, |t1-t2|的值小到可以忽略.考虑到误差的干扰, 甚至可能遇见有索引效率更低的怪事.

做好这个实验的要点是:

1. 数据量越大越好, 至少大于内存, 防止数据库一次性读入全部数据. 因此, 你需要自己构造实验表格. 构造方法: 利用C++或者python代码的数据库接口, 随机且自动的生成海量数据, 每行数据包含一个1M左右的视频, 可以缩短构造时间. 但是视频不能太大, 例如1G, 因为这样表格行数会变少, 索引的价值就会降低.
2. 为了消除误差, 需要随机查询多次, 例如一万次, 将其平均值作为两种不同情况的时间开销. 这同样需要C++或者python代码操纵数据库.
3. 最好使用**硬盘是磁盘**这样的慢速存储的计算, 而非采用硬盘是固态硬盘的计算机. 存储设备相对于内存速度越慢, 对比越强烈.

教学在线中的文档“索引最终版”是往届同学的作品，共大家参考。考虑到工作量太大，我并不要求大家这样做。以往实验的结果显示, 无索引的开销是有索引的开销的十万倍到百万倍之间.

## 五、评分标准

在例行评分标准之外, 增加标准如下

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 是否根据老师的提示采用高级编程语言做本实验 | 现场 | 现场得分 | 文档 | 文档得分 |
| 否 | 简单输入SQL命令,比较两种情况的时间. 能完美解释, 为什么用了索引更慢?为什么用不用差不多? | 89 | 能完美解释, 为什么用了索引更慢?为什么用不用索引差不多? | 89 |
| 否 | 简单输入SQL命令,比较两种情况的时间. 无法解释, 为什么用了索引更慢?为什么用不用差不多? | 80 | 无法解释, 为什么用了索引更慢?为什么用不用索引差不多? | 80 |
| 是 | 1. 实现了要点1. 由于代码生成数据需要很长时间, 不要求现场生成数据. 但需要现场运行相关代码, 证明确实能自动生成数据即可. 2. 实现了要点2. 这个过程时间开销不大, 需要现场展示和运行相关代码. 3. 口头能对相关代码做介绍和说明. | 100 | 文档包含如下内容   1. 设计思路 2. 关键代码模块关系图 3. 回答:没有索引的开销是有索引开销的多少倍 | 100 |
| 是 | 上述要点, 个别地方做得有缺陷 | 适当减分 | 上述要点, 个别地方做得有缺陷 | 适当减分 |
| 是 | 没有做到上述任何要点 | 80 | 没有做到上述任何要点 | 80 |