

数据库系统概论中期报告

刘轩奇 2018011025

2020.11.16

1 小组分工

本次数据库大作业由本人一人完成。

2 进展汇报

2.1 页式文件系统

课程提供的页式文件系统使用起来并不顺畅，尤其是在多文件处理上有问题。具体而言，同时打开多个文件时若稍有不慎则容易使得多个文件分配同一个文件编号。因而最终使用了网上找到的另一个页式文件系统代码¹。

2.2 记录管理模块

使用单个文件维护单个表。文件的第一页为元信息，记录表中已有的记录数、属性的数量和类型等。从第二页开始，每页记录各条记录。已经支持的功能有：

1. 定义整数、双精度浮点数、定长字符串类型的属性。
2. 创建和加载表文件。
3. 插入、删除和修改记录。
4. 遍历全表查询值符合某条件的记录。
5. 在一次查询结束后的结果中再次指定新条件查询。
6. 多个查询求交、求并、求差。

2.3 索引管理模块

使用一个文件维护一个索引。文件第一页为元信息，记录索引的键域、键域的类型信息等。从第二页开始，每页维护一个B+树的节点。节点分为内部节点和叶子节点。内部节点存储每个子节点的页面编号，以及该子页面的最小键值。每个叶子节点存储一条记录在记录文件中的位置（用页面号和页面中的记录号表示），以及该记录的索引键值。已经支持的功能有：

¹<https://github.com/Harry-Chen/SimpleDB/tree/master/src/io>

1. 创建单个键值的索引，以及创建联合索引。
2. 逐条向索引中添加记录，页面填满时自动分裂。
3. 绑定或解绑数据表。
4. 添加多条记录后重新创建索引。
5. 根据键值的有序性高效查询。包括闭区间查询、开区间、半开区间查询等。
6. 查询结果求交、求并、求差。

3 问题和困难

1 类型处理 支持整数、浮点数、定长字符串类型（字符类型可以看作长度为2的字符串类型）的处理时（例如比较、插入、修改）是分别使用不同函数处理的。这种不一致性容易导致难以维护。

2 页式文件 由于不知道页式文件系统系统会何时舍弃旧页面，因而每次想要读取文件的元数据页面时都需要重新确保加载该页，可能导致 IO 次数增加，效率变慢。

3 文件读取 在本人的 Window 10 x64 机器上，文件写入和读取似乎并不对称。写入 8196 个字节的数据，关闭文件，再从同一个文件的同一个位置读取这 8192 个字节，则重新读出来的数据与写入的数据有不同。但是在 Ubuntu 20.04 Linux x64 系统上则没有此问题。

4 建议

本课程提供的页式文件系统有时会出现意料不到的错误，如上文所述。希望未来提供页式文件系统时能够确保提供代码的可用性。