|  |
| --- |
| 哈尔滨工业大学(深圳) |
| **《数据库》实验报告** |
|  |
| **实验五**  **查询处理算法的模拟实现**  学 院: 计算机科学与技术   |  |  | | --- | --- | | 姓 名: | 邓佳淇 | | 学 号: | 200110717 | | 专 业: | 计算机科学与技术 | | 日 期: | 2022-01-04 | |

# 实验目的

*阐述本次实验的目的。*

理解索引、散列的作用

掌握关系选择、投影、连接、集合的交、并、差等操作的实现算法

加深对算法I/O复杂性的理解。

# 实验环境

*阐述本次实验的环境。*

macOS

Xcode

# 实验内容

*阐述本次实验的具体内容。*

基于ExtMem程序库，实现关系选择、连接操作算法。

实现集合并、交、差操作算法

# 实验过程

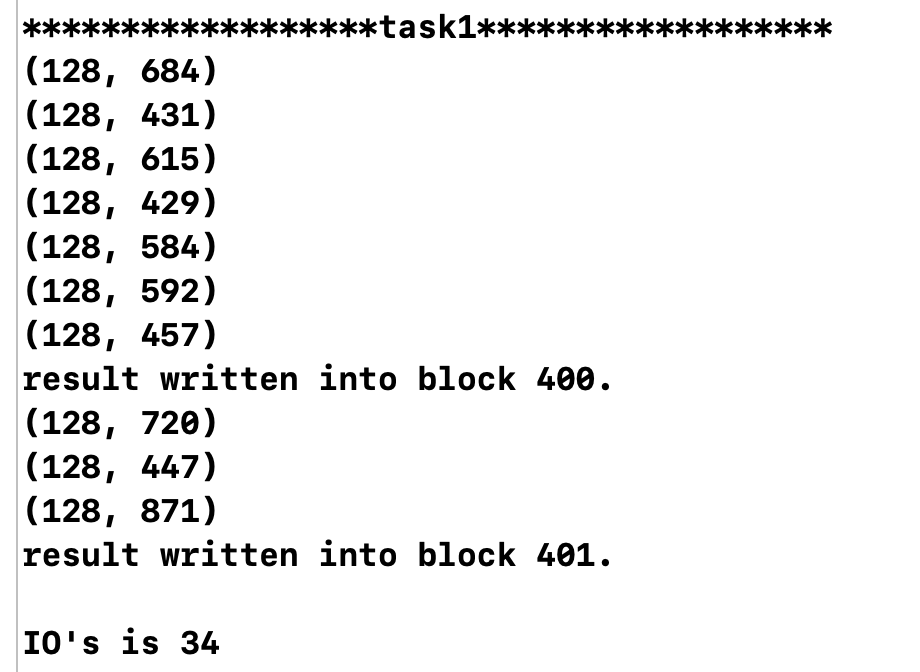
*对实验中的5个题目分别进行分析，并对核心代码和算法流程进行讲解，用自然语言描述解决问题的方案 。并给出程序正确运行的结果截图。*

1. **实现基于线性搜索的关系选择算法**

问题分析：线性搜索即遍历集合找出符合条件的数据

对S集合进行遍历，将磁盘上存储的S集合的数据读到内存中，找到C属性为128的元组，写入磁盘。

实验结果：

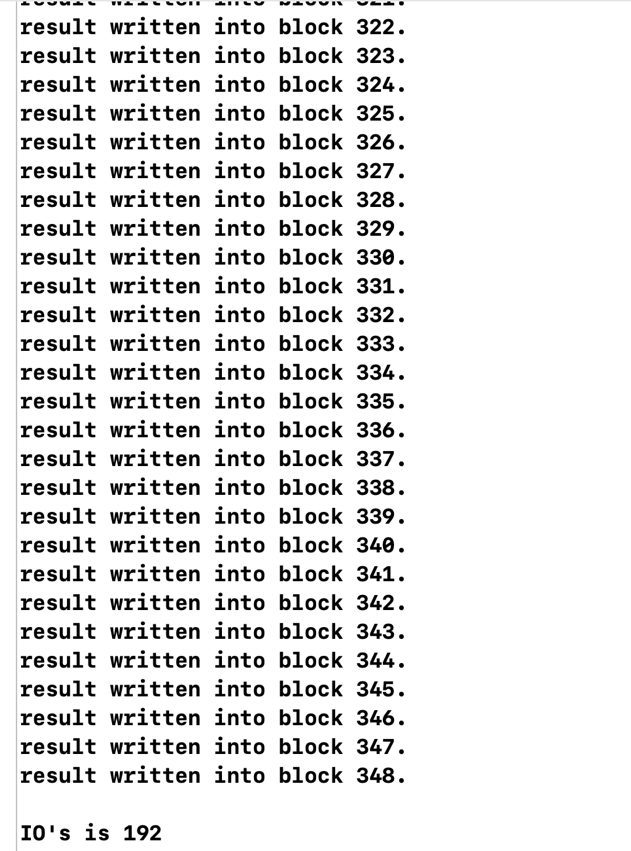
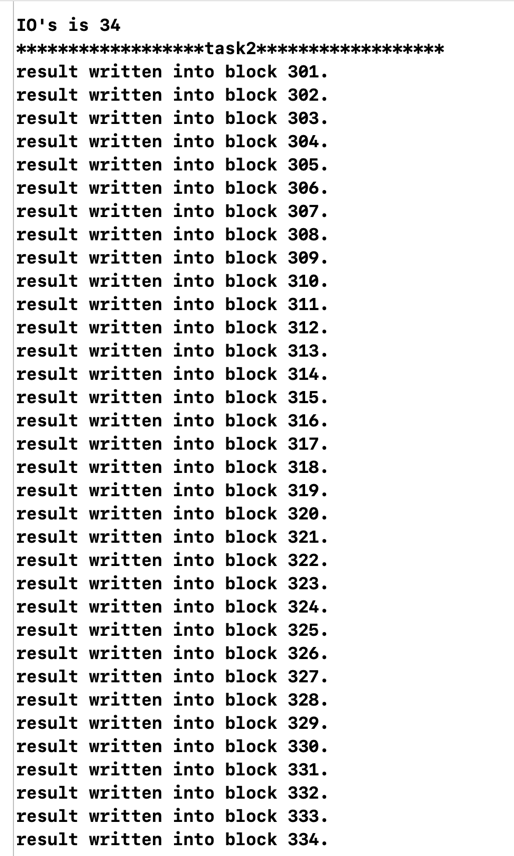


1. **实现两阶段多路归并排序算法（TPMMS）**

问题分析：

分为两阶段：第一阶段划分子集并将子集排序。第二阶段各个子集之间归并排序。第一阶段：划分为7块一个子集。一次性将一个子集读入到内存，并进行排序，写回磁盘.第二阶段：将排好序的子集分别读入内存两块，并对各个子集同时进行比较，将最小的写入内存中的空闲块，当空闲块写满时，将其写回磁盘。当集合的块被比较完成后，继续读入该集合的另一块直到所有块都比较完成。

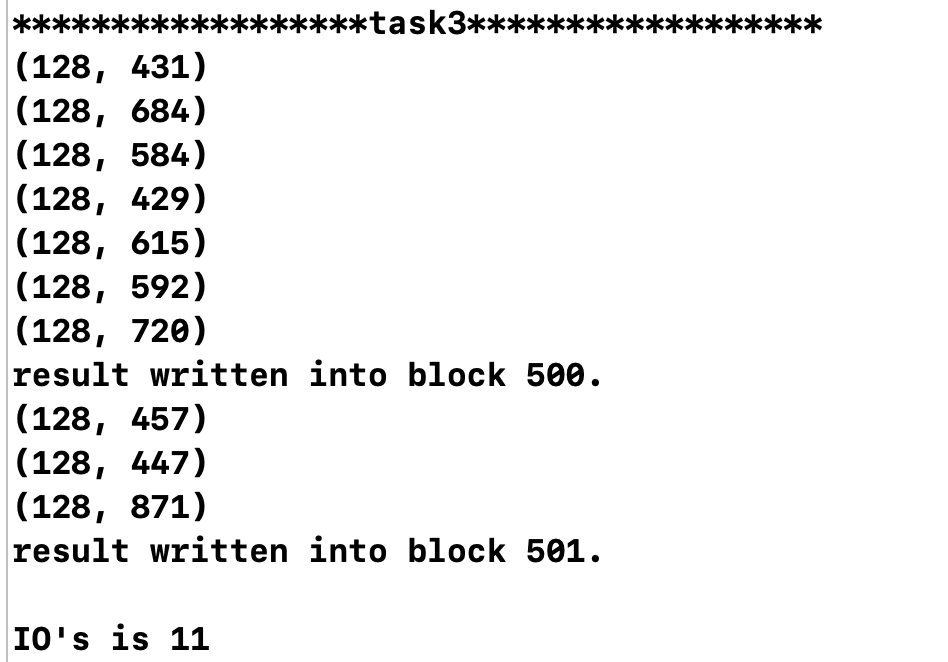
实验结果：



1. **实现基于索引的关系选择算法**

问题分析：读入排好序的数据，从第一个数据开始与要查找的数据进行比较，若不同则继续比较下一个数据，若相同，则将符合条件的元组全部写入结果磁盘。

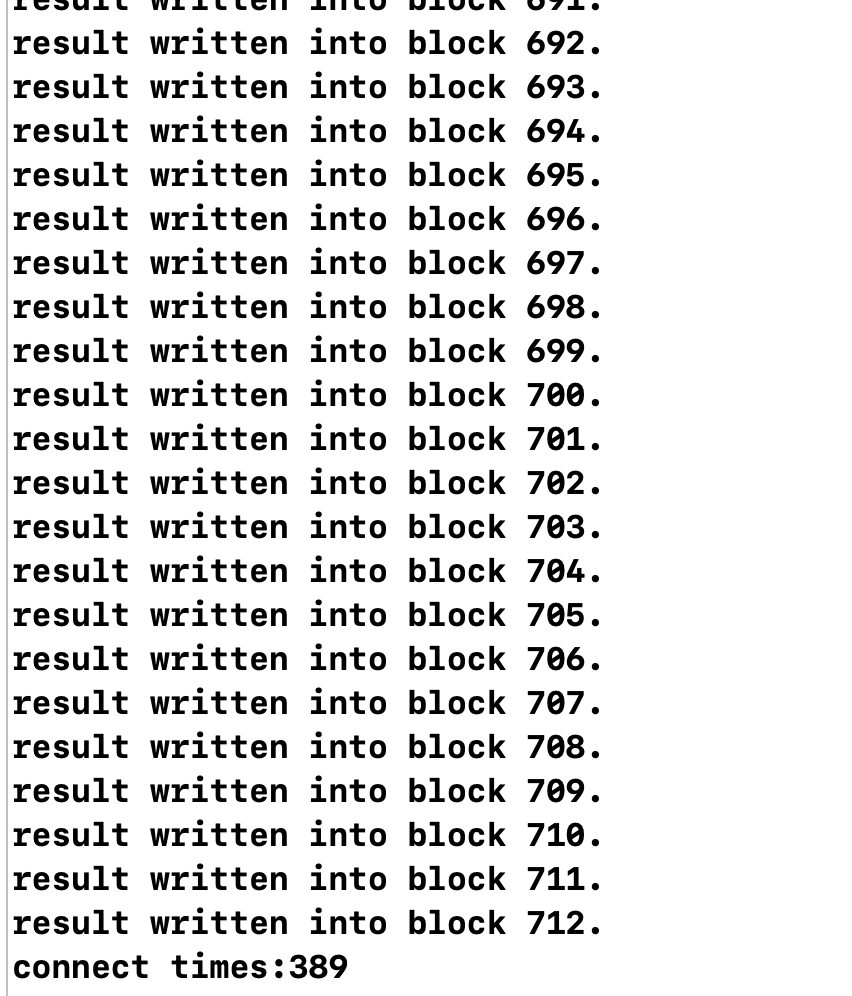
实验结果：



1. **实现基于排序的连接操作算法（Sort-Merge-Join）**

问题分析：将排好序的两个集合的数据分别读入内存，比较两个集合的数据的第一个属性。若不相同，则将拥有较小数据的集合的指针增加；若相同，则保持S集合指针不动，R集合临时指针不断向前移动，直到两个数据不相等。将过程中相等的元组对写入结果磁盘块。恢复R集合指针，并将S集合指针增加。重复上述过程直到指针到达两集合的最后。

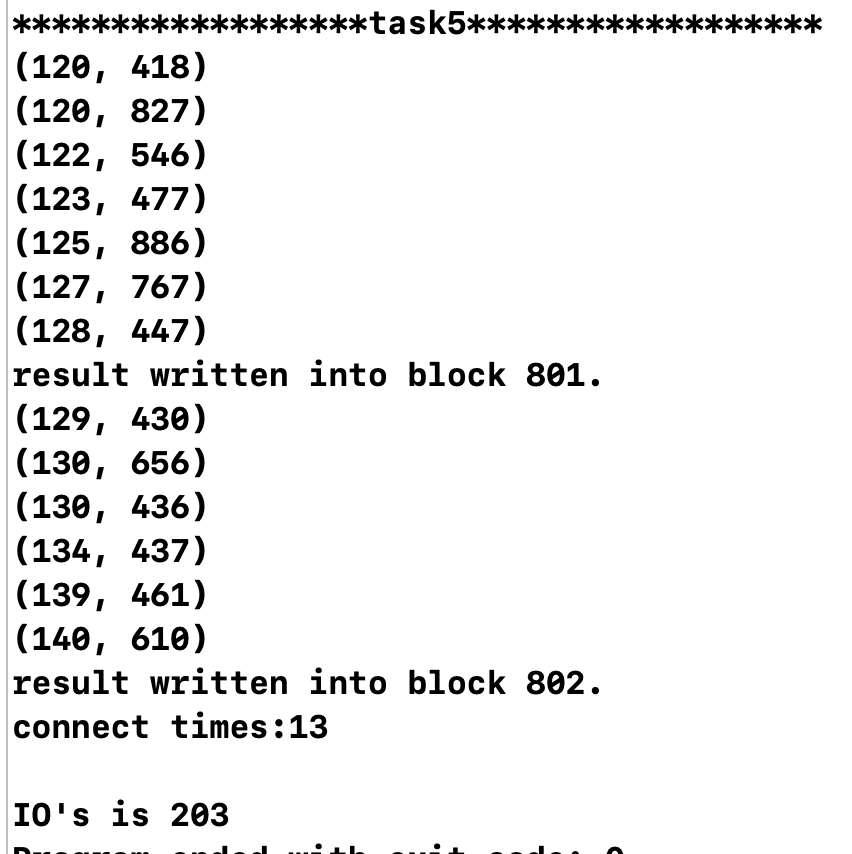
实验结果：



1. **实现基于散列的两趟扫描算法，实现交、并、差其中一种集合操作算法**

问题分析：将排好序的两个集合的数据分别读入内存，比较两个集合的数据。若不相同，则将拥有较小数据的集合的指针增加；若相同，则保持S集合指针不动，R集合临时指针不断向前移动，直到两个数据不相等。将过程中相等的元组对写入结果磁盘块。恢复R集合指针，并将S集合指针增加。重复上述过程直到指针到达两集合的最后。

实验结果：



# 附加题

*对剩余的两种集合操作进行问题分析，并给出程序正确运行的结果截图。*

# 总结

*总结本次实验的遇到并解决的问题、收获及反思。*

进行内存访问时，地址很容易混乱。需要建立单独的函数进行内存的读写。