|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 |  | 成  绩 |  |
| 学号 |  |

第七章 作业

【题目】

1. 简单比较文件的多重表和倒排表组织方式各有什么优缺点。
2. 假设原有一个计算机系统，能够处理一个大公司所有雇员记录的排序操作。由于公司合并，其雇员数目扩大为原来的100倍。计算机在可接受的时间内不能对原来100倍的记录进行外排序了。下面这几种独立的方案考虑对计算机系统进行升级，它们是否能够满足新的外排序要求？
3. CPU的速度增长两倍；
4. 磁盘I/O时间缩短至原来的一半；
5. 内存访问时间缩短至原来的一半；
6. 内存大小增长两倍；
7. 如果某个文件经内排序得到80个初始归并段，试问：
8. 若使用多路归并执行3 趟完成排序，那么应取的归并路数至少应为多少？
9. 如果操作系统要求一个程序同时可用的输入/输出文件的总数不超过15个，则按多路归并至少需要几趟可以完成排序？如果限定这个趟数，可取的最低路数是多少？

【答题】（宋体小四号，单倍行距）

1、解答：

**多重表文件：记录按主关键字的顺序够成一个串联文件，并建立主关键字的索引；对每一个次关键字项建立次关键字索引。所有具有同一关键字的记录够成一个链表。**

**优点：易于构造，也易于修改。如果不要求保持链表的某种次序，则插入记录容易，此时可将记录插在链表的头指针之后。**

**缺点：删除一个记录较繁琐，需要在每个次关键字的链表中删去该记录。**

**倒排文件：用记录的非主属性值（也叫副键）来查找记录而组织的文件**

**优点：检索记录较快**

**缺点：维护困难，在同一索引表中，不同关键字记录数不同，各倒排表的长度不等，同一倒排表中各项长度也不等。**

2、解答

外排序的主要时间花销在于磁盘访问的I/O时间，内存中进行的内排序的时间可以在磁盘访问的同时进行，因此内排序的时间可以忽略不计。因此：

1. CPU的速度增长两倍并不能明显加快外排序的速度；
2. 磁盘I/O时间缩短至原来的一半可以将外排序速度提高一倍；
3. 内存访问时间缩短至原来的一半并不能加快外排序的速度；
4. 内存大小增长两倍虽然可以增加内排序的规模（变成原来的两倍），但是这也需要更多的数据来填充数据缓冲区，因此依然不能明显加快外排序的速度。

3、解答

1. 设归并路数为k，k3≥80，由此解得k≥5，即应取的归并路数至少为5。
2. 设多路归并的归并路数为k，需要k个输入缓冲区和1个输出缓冲区，k+1=15，因此k=14，可做14路归并。由S=log1480=2，即至少需2趟归并可完成排序。

若限定这个趟数，有80≤k2，可取的最低路数为9。

对于所谓“趟”的理解如下：

