第十一章 文件

一．选择题

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.D | 2.A | 3.B | 4.A | 5.B | 6.B | 7.B |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

二.判断题

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.√ | 2.√ | 3.× | 4.√ | 5.× | 6.× | 7.× | 8.× | 9.× | 10.× | 11.√ |  |

三．填空题

1．操作系统文件 数据库 2．单关键字文件 多关键字文件

3．（1）数据库 （2）文本 （3）顺序组织 （4）随机组织 （5）链组织

（6）随机组织 （7）m （8）⎡m/2⎤ （9）2 （10） k

4．记录 数据项 5．串联文件 6．第I-1 7．随机

8．提高查找速度 9．树 10．检索记录快

11．（1）关键字 （2）记录号 （3）记录号 （4）顺序 （5）直接

12．构造散列函数 解决冲突的方法 13．索引集 顺序集 数据集

14．分配和释放存储空间 重组 对插入的记录

四．应用题

1．文件是由大量性质相同的记录组成的集合，按记录类型不同可分为操作系统文件和数据库文件。

2．文件的基本组织方式有顺序组织、索引组织、散列组织和链组织。文件的存储结构可以采用将基本组织结合的方法，常用的结构有顺序结构、索引结构、散列结构。

（1） 顺序结构，相应文件为顺序文件，其记录按存入文件的先后次序顺序存放。顺序文件本质上就是顺序表。若逻辑上相邻的两个记录在存储位置上相邻，则为连续文件；若记录之间以指针相链接，则称为串联文件。顺序文件只能顺序存取，要更新某个记录，必须复制整个文件。顺序文件连续存取的速度快，主要适用于顺序存取，批量修改的情况。

（2） 带索引的结构，相应文件为索引文件。索引文件包括索引表和数据表，索引表中的索引项包括数据表中数据的关键字和相应地址，索引表有序，其物理顺序体现了文件的逻辑次序，实现了文件的线性结构。索引文件只能是磁盘文件，既能顺序存取，又能隋机存取。

（3） 散列结构，也称计算寻址结构，相应文件称为散列文件，其记录是根据关键字值经散列函数计算确定其地址，存取速度快，不需索引，节省存储空间。不能顺序存取，只能随机存取。

其它文件均由以上文件派生而得。

文件采用何种存储结构应综合考虑各种因素，如：存储介质类型、记录的类型、大小和关键字的数目以及对文件作何种操作。

3．在主文件外，再建立索引表指示关键字及其物理记录的地址间一一对应关系。这种由索引表和主文件一起构成的文件称为索引文件。索引表依关键字有序。主文件若按关键字有序称为索引顺序文件，否则称为索引非顺序文件（通常简称索引文件）。索引顺序文件因主文件有序，一般用稀疏索引，占用空间较少。常用索引顺序文件有ISAM和VSAM。ISAM采用静态索引结构，而VSAM采用B+树的动态索引结构。索引文件既能顺序存取，也能随机存取。

4．在索引文件中，若（数据区）主文件中关键字有序，则文件称为索引顺序文件，参见上题3。

5．ISAM是专为磁盘存取设计的文件组织方式。即使主文件关键字有序，但因磁盘是以盘组、柱面和磁道（盘面）三级地址存取的设备，因此通常对磁盘上的数据文件建立盘组、柱面和磁道（盘面）三级索引。在ISAM文件上检索记录时，先从主索引（柱面索引的索引）找到相应柱面索引。再从柱面索引找到记录所在柱面的磁道索引，最后从磁道索引找到记录所在磁道的第一个记录的位置，由此出发在该磁道上进行顺序查找直到查到为止；反之，若找遍该磁道而未找到所查记录，则文件中无此记录。

6．ISAM是一种专为磁盘存取设计的文件组织形式，采用静态索引结构，对磁盘上的数据文件建立盘组、柱面、磁道三级索引。ISAM文件中记录按关键字顺序存放，插入记录时需移动记录并将同一磁道上最后的一个记录移至溢出区，同时修改磁道索引项，删除记录只需在存储位置作标记，不需移动记录和修改指针。经过多次插入和删除记录后，文件结构变得不合理，需周期整理ISAM文件。

VSAM文件采用B+树动态索引结构，文件只有控制区间和控制区域等逻辑存储单位，与外存储器中柱面、磁道等具体存储单位没有必然联系。VSAM文件结构包括索引集、顺序集和数据集三部分，记录存于数据集中，顺序集和索引集构成B+树，作为文件的索引部分可实现顺链查找和从根结点开始的随机查找。

与ISAM文件相比，VSAM文件有如下优点：动态分配和释放存储空间，不需对文件进行重组；能保持较高的查找效率，且查找先后插入记录所需时间相同。因此，基于B+树的VSAM文件通常作为大型索引顺序文件的标准组织。

7．ISAM文件有三级索引：磁盘组、柱面和磁盘，柱面索引存放在某个柱面上，若柱面索引较大，占多个磁道时，可建立柱面索引的索引—主索引。故本题中所指的两级索引是盘组和磁道。

8．倒排文件是一种多关键字的文件，主数据文件按关键字顺序构成串联文件，并建立主关键字索引。对次关键字也建立索引，该索引称为倒排表。倒排表包括两项，一项是次关键字，另一项是具有同一次关键字值的记录的物理记录号（若数据文件非串联文件，而是索引顺序文件—如ISAM，则倒排表中存放记录的主关键字而不是物理记录号）。倒排表作索引的优点是索引记录快，缺点是维护困难。在同一索引表中，不同的关键字其记录数不同，各倒排表的长度不同，同一倒排表中各项长度也不相等。

9．因倒排文件组织中，倒排表有关键字值及同一关键字值的记录的所有物理记录号，可方便地查询具有同一关键字值的所有记录；而多重表文件中次关键字索引结构不同，删除关键字域后查询性能受到影响。

10．多重表文件是把索引与链接结合而形成的组织方式。记录按主关键字顺序构成一个串联文件，建立主关键字的索引（主索引）。对每一次关键字建立次关键字索引，具有同一关键字的记录构成一个链表。主索引为非稠密索引，次索引为稠密索引，每个索引项包括次关键字，头指针和链表长度。多重表文件易于编程，也易于插入，但删除繁锁。需在各次关键字链表中删除。倒排文件的特点见上面题8。

11．倒排表作索引的优点是索引记录快，因为从次关键字值直接找到各相关记录的物理记录号，倒排因此而得名（因通常的查询是从关键字查到记录）。在插入和删除记录时，倒排表随之修改，倒排表中具有相同次关键字的记录号是有序的。

12．排表有两项，一是次关键字值，二是具有相同次关键字值的物理记录号，这些记录号有序且顺序存储，不使用多重表中的指针链接，因而节省了空间。

13．（1）顺序文件只能顺序查找，优点是批量检索速度快，不适于单个记录的检索。顺序文件不能象顺序表那样插入、删除和修改，因文件中的记录不能象向量空间中的元素那样“移动”，只能通过复制整个文件实现上述操作。

（2）索引非顺序文件适合随机存取，不适合顺序存取，因主关键字未排序，若顺序存取会引起磁头频繁移动。索引顺序文件是最常用的文件组织，因主文件有序，既可顺序存取也可随机存取。索引非顺序文件是稠密索引，可以“预查找”，索引顺序文件是稀疏索引，不能“预查找”，但由于索引占空间较少，管理要求低，提高了索引的查找速度。

（3）散列文件也称直接存取文件，根据关键字的散列函数值和处理冲突的方法，将记录散列到外存上。这种文件组织只适用于像磁盘那样的直接存取设备，其优点是文件随机存放，记录不必排序，插入、删除方便，存取速度快，无需索引区，节省存储空间。缺点是散列文件不能顺序存取，且只限于简单查询。经多次插入、删除后，文件结构不合理，需重组文件，这很费时。

14．类似最优二叉树（哈夫曼树），可先合并含较少记录的文件，后合并较多记录的文件，使移动次数减少。见下面的哈夫曼树。

|  |
| --- |
| F2A |

|  |
| --- |
| F5 |

|  |
| --- |
|  |

|  |
| --- |
| F3 |

|  |
| --- |
|  |

|  |
| --- |
| F4 |

|  |
| --- |
| F1 |

|  |
| --- |
|  |

|  |
| --- |
|  |

15．[问题分析]在职务项中增加一个指针项，指向其领导者。因题目中未提出具体的隶属关系，如哪个系的系主任，哪个系哪个室的室主任，哪个室的教员等。这里假设每个室主任隶属于他前边离他最近的那个系主任，每个教员隶属于他前边离他最近的那个室主任，见下面多重表文件。在职称项中增加一个指针项，指向同一职称的下一个职工，增加一个次关键字索引表：

|  |  |  |
| --- | --- | --- |
| 关键字 | 头指针 | 长度 |
| 讲师 | 001 | 2 |
| 副教授 | 004 | 2 |
| 教授 | 002 | 6 |

“职称”索引表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 记录号 | 职工号 | 职工姓名 | 职务 | | 职称 | |
| 01  02  03  04  05  06  07  08  09  10 | 001  002  003  004  005  006  007  008  009  010  … | 张军  沈灵  叶明  张莲  叶宏  周芳  刘光  黄兵  李民  赵松  … | 教员  系主任  校长  室主任  系主任  教员  系主任  教员  室主任  教员  … | 04  03  ∧  02  03  04  03  04  07  09 | 讲师  教授  教授  副教授  教授  教授  教授  讲师  教授  副教授  … | 08  03  05  10  06  07  09  ∧  ∧  ∧ |
|  |  |  |  |  |  |  |

多重表文件