考研 408—操作系统讲义【文件管理】

一. 文件和文件系统

文件系统的管理功能是将其管理的程序和数据通过组织为一系列文件的方式实现

- 1. 数据项 记录、文件
 - I. 数据项

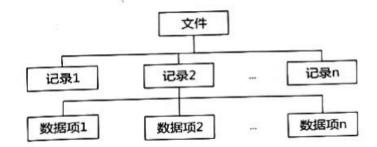
数据项是文件系统中最低级的数据组织形式,可分为以下两种类型:

- i) 基本数据项:用于描述一个对象的某种属性的一个值,如姓名、日期或证件号等, 是数据中可命名的最小逻辑数据单位,即原子数据
- ii) 组合数据项:由多个基本数据项组成。
- II. 记录

记录是一组相关的数据项的集合,用于描述一个对象在某方面的属性,如一个考生报名记录包括考生姓名、出生日期、报考学校代号、身份证号等一系列域

III. 文件

① 文件类型 ② 文件长度 ③ 文件的物理位置 ④ 文件的建立时间



2. 文件名和类型

I. 文件名和扩展名

在不同的系统中,对文件名的规定是不同的,在一些老的系统中,名字的长度受到 限制。

扩展名是添加在文件后面的若干附加字符,有称为后缀名,用于指示文件的类型。

II. 文件类型

按文件的性质和用途分类

- i) 系统文件:系统软件构成的文件,如内核,系统应用程序,数据;只允许用户执行, 不能读写和修改
- ii) 用户文件:由用户建立的文件,如源程序、目标程序和数据文件等;用户将这些文件托给系统保管;只允许文件所有者和所有者授权用户使用
- iii) 库文件:标准子例程及常用的例程构成的文件。这类文件允许用户调用,但不允许用户修改

按文件中数据的形式分类

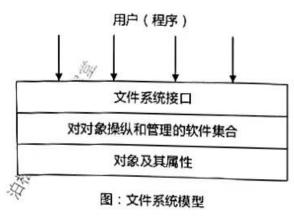
- i) 源文件:由源程序和数据构成的文件。通常由终端或输入设备输入的源程序和数据 所形成的文件都属于源文件
- ii)、目标文件: 把源程序经过相应语言的编译程序编译过, 但尚未经过链接程序链接的目标代码所构成的文件。它属于二进制文件。通常, 目标文件所使用的后缀名是 ".obj"
- iii) 可执行文件:把编译后所产生的目标代码再经过链接程序链接后所形成的文件 按**存取控制属性分类**
- i) 只执行文件:只允许被核准的用户调用执行,既不允许读,更不允许写
- ii)只读文件:只允许文件主及被核准的用户去读,但不允许写
- iii)读写文件:允许文件主和被核准的用户去读或写的文件
- iv)无保护文件

按文件的内容形式和系统处理方式分类。

- i) 普通文件:由 ASCII 码或二进制码组成的文件。一般用户建立的源程序文件、数据文件、目标代码文件及操作系统自身代码文件、库文件等都是普通文件,它们通常存储在外存储设备上
- ii) 目录文件:由文件目录组成的,用来管理和实现文件系统功能的系统文件,通过目录文件可对其它文件的信息进行检索。目录文件也是由字符序列构成,可进行与普通文件一样的各种操作
- iii) 特殊文件(设备文件):特指系统中的各类 I/O 设备。为了便于统一管理,系统将所有的输入/输出设备都视为文件,按文件方式提供给用户使用。

3. 文件系统的层次结构

文件系统的模型可以分为三个层次:最底层是对象及其属性,中间层是对对象进行操纵和管理的软件集合,最高层是文件系统提供给用户的接口





I. 对象及其属性

文件管理系统管理的对象如下:

- ① 文件 ② 目录 ③ 磁盘存储空间
- II. 对对象操作和管理的软件集合

文件系统的功能大多是在这一层实现的,其中包括有:

- i) 对文件存储空间的管理
- ii)》对文件目录的管理
- iii) 用于将文件的逻辑地址转换为物理地址的机制
- iv) 对文件读和写的管理
- v) 对文件的共享与保护等功能

把与文件系统有关的软件分为四个层次

- i) I/O 控制层: 文件系统的最低层, 主要由磁盘驱动程序等组成, 称为设备驱动程序
- ii) 基本文件系统层:主要用于处理内存与磁盘之间数据块的交换
- iii) 基本 I/O 管理程序:该层用于完成与磁盘 I/O 有关的事务
- iv)逻辑文件系统:用于处理与记录和文件相关的操作

文件系统的接口

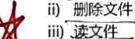
- i) 命令接口:是指作为用户与文件系统交互的接口,用户可通过键盘终端键入命令取 得文件系统的服务
- ii) 程序接口:是指作为用户程序与文件系统的接口,用户程序可通过系统调用取得文 件系统的服务

4. 文件操作

最基本的文件操作包括创建、删除、读、写和设置文件的读/写位置 一般 OS 都提供了更多对文件的操作,如打开和关闭一个文件及改变文件名等等

I. 最基本的操作

i) 创建文件



ii) 删除文件

iv) 写文件_

v) 设置文件的读/写的位置(才ど学十)

II. 文件的打开和关闭操作

当用户要求对一个文件实施多次读/写操作时/每次都要从检索目录开始,为了避 免多次重复检索目录,在大多数 OS 中引入了"打开"这一女件系统调用,当用户第一次 请求对某文件进行操作时,须先利用 open 系统调用将该文件打开。

III. 其他文件操作

常用的一类是有关对文件属性的操作,即允许用户直接设置和获取文件的属性

二. 文件的逻辑结构

用户看到的文件称为逻辑文件,它是由一系列的逻辑记录组成的。文件的逻辑记录时能够



被存取的基本单位

系统中的所有文件都存在两种形式的文件结构:

I. 文件的逻辑结构:从用户观点出发所观察到的文件组织形式

II. 文件的物理结构:文件的存储结构

1. 文件逻辑结构的类型

X

I. 按文件是否有结构分类

- i) 有结构文件ン
 - √ 定长记录:文件中所有记录的长度都是相同的,个数据项都处在记录中相同的 位置,具有相同的顺序和长度。

优点:有效提高检索记录的速度和效率,方便对文件进行处理和修改

◆ 变长记录:文件中各记录的长度不相同,产生边长记录的原因可能是由于一个 记录中所包含数据项数目并不相同。

缺点:记录的检索速度慢,不方便对文件进行处理和修改

ii)无结构文件)

系统中运行的大量的源程序、可执行文件、库函数等采用的是无结构文件形式 (流式文件)・ としている。

II. 按文件的组织方式分类

i) 顺序文件 指由一系列记录按某种顺序排列所形成的文件,其记录可以是定长记录 也选长记录

→ ii) 索引文件,指为可变长记录文件建立一张索引表,为每个记录设置一个表项,以加快对记录的检索速度。

文 iii)索引顺序文件为一个文件建立一张索引表,并不是为每一个记录建立一个索引表项。

- 2. 顺序文件
 - 顺序文件的排列方式
 - i) 串结构

按照存入时间的先后进行排序,各个记录之间的顺序与关键字无关。对文件进行检索的时候,每次都要从头开始,逐个记录的查找。缺点:耗时长

ii) 顺序结构

用户指定哪一个字段作为关键字。优点:提高检索效率

II. 顺序文件的优缺点

优点: ① 存取效率高

缺点:① 查找性能差 ② 增加删除困难

- 3. 记录寻址
 - I. 隐式寻址方式

对于定长的顺序文件,如果已知当前记录的逻辑地址,很容易确定下一个记录的逻辑地址。

II. 显式寻址方式

用于对定长记录的文件实现直接或随机访问

- i) 通过文件中记录的位置
- ii) 关键字

4. 索引文件

I. 按关键字建立索引

索引号	长度m	指针ptr		R ₀
0 (m_0			R ₁
1	m ₁			i
, :	:		/	Ri
i	mi			:
:	i			逻辑文件

索引表

图: 具有单个索引表的索引文件



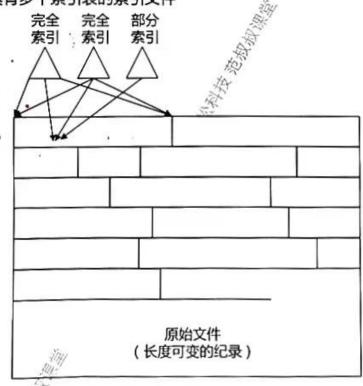


图: 具有多个索引表的索引文件

5. 索引顺序文件

I. 索引顺序文件的特征

优点:克服变长记录的顺序文件不能随机访问,以及不便记录的删除和插入的缺点两个新特征;① 引入文件索引表 ② 增加溢出文件

II. 一级索引顺序文件

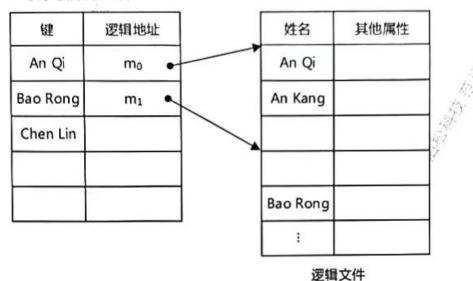


图:索引顺序文件

III. 两级索引文件

6. 直接文件和哈希文件

I. 直接文件

对于直接文件,则根据给定的记录键值,直接获得指定记录的物理地址,换言之, 记录键值本身就决定了记录的物理地址,这种由记录键值到记录物理地址的转换被称 为键值转换。

II. 哈希文件



利用 Hash 函数可将记录键值转换为相应记录的地址,为了能实现文件存储空间的动态分配,通常由 Hash 函数所求得的并非是相应记录的地址,而是指向一目录表相应表目的指针,该表目的内容指向相应记录所在的物理块。

三、文件目录一态和的了文件FCB· FCBT与介序原含

为了能够对文件实施有效的管理,必须对它们加以妥善组织,这主要是通过文件目录实现的,**文件目录也是一种数据结构,用于标识系统中的文件及其物理地址,供检索时使用**,对目录的管理要求如下:

- <u>实现按名存取</u>:即用户只须向系统提供所需访问的文件的名字,便能够快速准确地找到指 定文件在外存上的存储位置,这是目录管理中最基本的功能
- 提高对目录检索速度:通过合理地组织目录结构的方法,可加快对目录的检索速度,从而提高对文件的存取速度
- 文件共享: 在多用户系统中, 应该允许用户共享一个文件
- ◆ 允许文件重名: 系统应允许不同用户对不同文件采用相同的名字,以便用户按照自己的习惯给文件命名和使用文件
- 1. 文件控制块

为了能对文件进行正确的存取,必须为文件设置用于描述和控制文件的数据结构, 称之为文件控制块 FCB,文件管理程序可借助于文件控制块中的信息,对文件施加各种操作,



分

考研 408——操作系统讲义【文件管理】

文件与文件控制块一对应,而人们把文件控制块的有序集合称为文件目录,一个文件控制块就是一个文件目录项。通常,一个文件目录也可被看成是一个文件,称为目录文件文件控制块包含基本信息、存取控制信息、使用信息

I. 基本信息类

- i) 文件名(标识一个文件的符号名,在每个系统中,每个文件都有唯一的名字,用户 利用该名字进行存取)
- ii) 文件物理位置(指文件在外存上的存储位置,包括存放文件的设备名、文件在外村上的起始盘块号、指示文件所占用的盘块数或字节数的文件长度)
- iii) 文件逻辑结构(指示文件是流式文件还是记录式文件、记录数,文件是定长还是变长记录)
- iv) 文件物理结构 (指示文件是顺序文件、链式文件还是索引文件)

Ⅱ. 存取控制信息

包括文件主的存取权限、核准用户的存取权限及一般用户的存取权限

Ⅲ. 使用信息

包括文件的建立日期和时间、文件上一次修改的日期和时间及当前使用信息(这项信息包括当前已打开该文件的进程数、是否被其他进程锁住、文件在内存中是否已被修改但尚未拷贝到盘上)

2. 索引结点

I. 索引结点的引入

文件目录通常是存放在磁盘上的,当文件很多时,文件目录可能要占用大量的盘块,在查找的过程中,先将存放目录文件的第一个盘块中的目录调入内存,然后把用户所给定的文件名和目录项中的文件名逐一对比。若未找到指定文件,则再将下一个盘块中的目录项调入内存。

在检索目录文件时,只用到了文件名,仅当找到一个目录项(即其中的文件名与指定要查找的文件名相匹配)时,才需要从该目录项中读出该文件的物理地址,而其他一些对该文件进行描述的信息,在检索目录时一概不用,显然,这些信息在检索目录时不需要调入内存。为此,在有的系统中,如UNIX系统,便采用了把文件名和文件描述信息分开的方法,亦即,使文件描述信息单独形成一个称为索引结点的数据结构,简称为i结点,在文件目录中的每个目录项由文件名和指向该文件所对应的i结点的指针所构成。

II. 磁盘索引结点

每个文件都有唯一的磁盘索引结点(磁盘索引结点信息与文件名等信息一起构成了 FCB)

还包括;

- i) 文件主标识符
- ii) 文件类型
- iii) 文件存取权限
- iv) 文件物理地址
- v) 文件长度
- vi) 文件连接计数

vii) 文件存取时间

(3)内存索引结点

当文件被打开,要将磁盘索引结点拷贝到内存的索引结点中,增加以下内容

- i) 索引结点编号
- ii) 状态 .
- iii) 访问计数
- iv)文件所属文件系统的逻辑设备号
- v) 链接指针

3. 目录结构

目录结构的组织,关系到文件系统的存取速度,也关系到文件的共享性和安全性,目前常用的目录结构形式有单级目录、两级目录、多级目录.

I. 单级目录结构

在整个系统中只建立一张目录表,每个文件占一个目录项,目录项中含文件名、文件扩展名、文件长度、文件类型、文件物理地址、状态位(表示目录项是否空闲)等每当要建立一个新文件时,必须先检查所有的目录项,以保证新文件名在目录中是唯一的,然后再从目录表中找到一个空白目录项,填入新文件的文件名及其他说明信息,并置状态为1,删除文件时,先从目录中找到该文件的目录项,回收该文件所占

用的存储空间,然后再清除该目录项

优点:是简单并且能够实现目录管理的基本功能-按名存取 缺点:

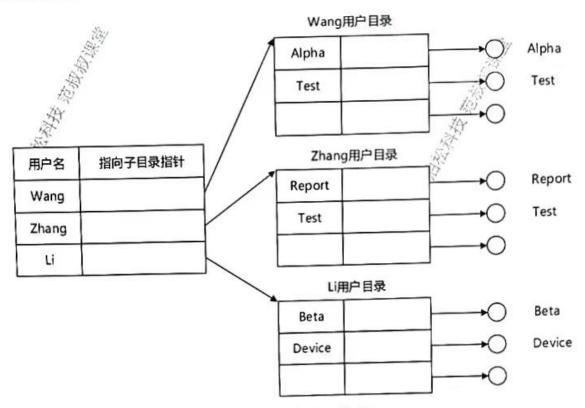
- 查找速度慢: 查找一个目录项要花费较多的时间
- 不允许重名:在一个目录表中的所有文件,都不能与另一个文件有相同的名字,这 是难以避免的
- 不便于实现文件共享:每一个用户都有自己的名字空间或命名习惯,因此,应该允许不同用户使用不同的文件名来访问同一个文件

II. 两级目录结构

每个用户建立一个单独的用户文件目录 UFD(User File Directory),这些文件目录具有相似的结构,由用户所有文件的文件控制块组成。此外,系统中还有一个主文件目录 MFD(Master File Directory),在主文件目录中,每个用户目录文件都占有一个目录项,其目录项包括用户名和指向用户目录文件的指针







两级目录结构克服了单级目录的缺点,具有如下优点:

- i) 提高了检索目录的速度:如果在主目录中有 n 个子目录,每个用户目录最多为 m 个目录项,则为查找一指定的目录项,最多只需要检索 n+m 个目录
- ii) 在不同的用户目录中,可以使用相同的文件名:只要在用户自己的 UFD 中,每个 文件名都是唯一的,不同用户可以有文件名相同的文件
- iii) 不同用户还可使用不同的文件名来访问系统中同一个共享文件:但在多个用户需要 合作完成一个大任务时,不便于用户之间共享文件

4. 树形目录结构

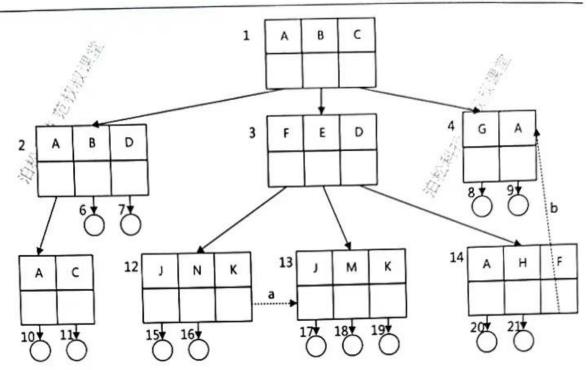
I. 树形目录

对于大型文件系统,通常采用三级或三级以上的目录结构,以提高对目录的检索速度和文件系统的性能。多级目录结构又称为树形目录结构,主目录被称为根目录,把数据文件称为树叶,其他的目录均作为树的结点

说明:方框代表目录文件,圆圈代表数据文件







Ⅱ. 路径名和当前目录

系统中的每个文件都有唯一的路径名

Ⅲ. 目录操作

- i) 创建目录
- ii) 删除目录
 - 不删除非空目录
 - 可删除非空目录
- iii) 改变目录
- iv) 移动目录
- v) 链接操作
- vi) 查找

5. 目录查询技术

当用户要访问一个已存在的文件时,系统首先利用用户提供的文件名对目录进行查询, 找出该文件的文件控制块或对应索引结点,然后,根据 FCB 或索引结点中所记录的文件物 理地址(盘块号),换算出文件在磁盘上的物理位置,最后,再通过磁盘驱动程序,将所 需文件读入内存。目前常用的方式有线性检索法和 Hash 方法

I. 线性检索法

其又称为顺序检索法,在单级目录中,利用用户提供的文件名,用顺序查找法从文件目录中找到文件的目录项,在树形目录中,用户提供的文件名是由多个文件分量名组成的路径名,此时须对多级目录进行查找,假定用户给定的文件路径名为/usr/ast/mbox,

考研 408——操作系统讲义【文件管理】

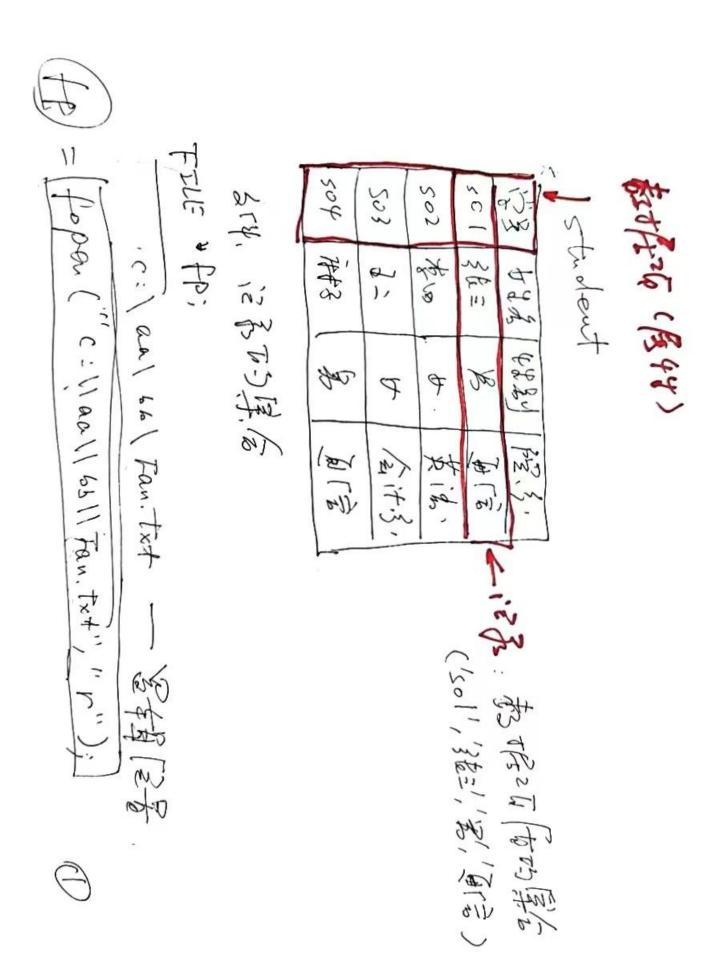
根目	录 频	节点6是/usr 的目录			节点26是 /usr/ast的目 录		盘块是 的目录
1			6			26	
1		132	1		496	6	
4	bin		19	dick	15	64	grants
7	dev		30	erik	12/5	92	books
14	lib		51	jim	AN AT	60	mbox
9	etc		26	ast		81	minik
6	usr		45	bal		17	src
	1 4 7 14 9	1	根目录 的目录 1	根目录 的目录 /usrß 1 .	根目录 的目录 /usr的目录 1 . 1 1 4 bin 7 dev 14 lib 9 etc 26 ast	根目录	村内

步骤:

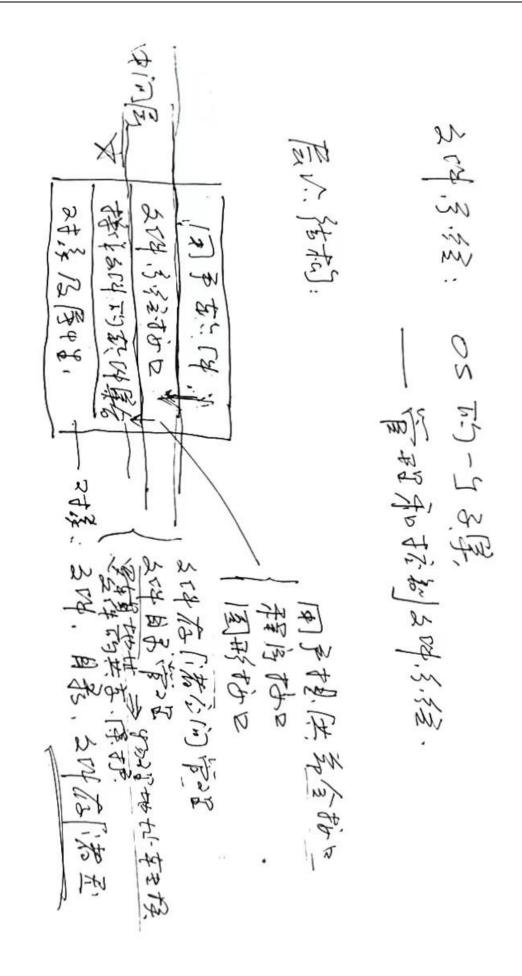
- i) 读入第一个文件分量名 usr,用它与根目录文件(或当前目录文件)中各目录项中的文件名顺序地进行比较
- ii) 得到匹配项的索引结点号是 6
- iii) 从 6 号索引结点中得到 usr 目录文件放在 132 号盘块中,将该盘块内容读入内存
- iv) 再将路径名中的第二个分量名 ast 读入,用它与放在 132 号盘块中的第二级目录 文件中各目录项的文件名顺序进行比较
- v) 依次类推

II. Hash 方法

系统利用用户提供的文件名并将它转换为文件目录的索引值,再利用该索引值到目录中去查找,这将提高检索速度



20月至十岁 有多好多的一多种美质展的星后 204 Bal. 文好的帮她地. [選輯務於 了了影響 公外方过一 下分支车的(基本) 文叶文本文: 15 多好な黄文本を



7352 12 O 214 FACE: 11>、多、多、多、多、多、多、多、多、多、多、多、多、 11> 在至14县墓事至一5月民场一方(东下08.13. 与响居出

E MAN 35 254 (3) 门门、林阳保下的作法分对下的影影的好好的分别在分间 12 作品料品中的形度活工中的品质的 1 13254 read web Xing Xing Clare

③ 才程: 于丁开台叫 本了开与时: (第2时下13 FCB (名时居中))间的村态 coper BADA write

演点网:系经通生内存中分对下的下的Carp图中的、探阳的神像局部 面的物质的各种各种的创新的

1300 1 松村、层沙群西、阳台中郡楞溪的内谷

