文件

常考知识点

UNIX相关知识	1. UNIX采用树型目录结构,文件信息存放在索引节点中,超级块是用来描述文件系统的 2. UNIX系统所有设备都被视为特殊的文件 3. UNIX每个目录文件中,默认有两个目录项,"."表示当前目录,""表示父目录
各种表、图	1. 打开文件表:存放已打开文件信息的表,将指名文件的属性从外存复制到内存,再使用该文件时直接返回索引 2. 位图:详见磁盘管理方法 3. 空闲盘链表:详见磁盘管理方法 4. 索引表:记录每个文件所存放的盘块地址 5. 系统调用表:一张由指向实现各种系统调用的内核函数的函数指针组成的表,该表可以基于系统调用函数进行索引,来定位函数地址,完成系统调用
文件性能相关	 与单个文件长度/大小有关的因素:文件块大小,地址项个数,间接地址索引的级数(索引节点总数与单个文件大小无关) 提高文件访问速度的方法:提前读,延迟写,为文件分配连续的簇,采用磁盘高速缓存、文件在磁带上采用连续存放方式;在硬盘上不采用连续存放方法;在内存上采用随机存放方法 连续结构的文件数据读最快,链式的方法不能随机存取,索引的方法占面积
文件打开相关	• 文件打开open是把FCB读到内存,不是把文件内容读到内存(即不会读入文件数据,只有用read的时候才会读数据) • open参数包含文件名 • read/write参数不包含文件名,而是open返回的文件索引
文件保护相关	防止文件受损的方法→备份;用于多用户之间的存取权限保护→存取控制矩阵方法
<u>其他</u>	系统级安全管理包括注册和登陆用户确定文件的逻辑结构,由操作系统设计者根据文件存储器的特性确定文件的物理结构,一旦确定,由操作系统管理

基本概念

文件是什么?	文件是以硬盘为载体的存储在计算机上的信息集合文件可以是文本文档,图片,程序用户进行的输入输出中,以文件为基本单位
文件由什么组成?	1. <u>一块存储空间</u> 2. <u>分类和索引的信息</u> 3. <u>关于访问权限的信息</u>
文件有什么特性?	1. <u>可以长期存储在硬盘中</u> 2. <u>允许可控制的进程间共享访问</u> 3. <u>能够被组织成复杂的结构</u>
文件的属性/元数据	・ 文件名; 文件类型・ <u>创建者; 所有者</u>・ 位置; 大小; 保护; 创建信息

文件的数据结构

目录项	相关概念	• 且录项/FCB = 用来记录文件的名字,索引节点指针以及其他目录项的层级关联			
	包含信息	• <u>存取控</u>	 基本信息:文件名;文件物理位置,文件逻辑结构,文件物理结构 存取控制信息:文件主或核准用户或一般用户的存取权限 使用信息:文件创立时间,上次修改时间 		
	概念	索引节将文件	方点是用来记录文件的元信息,是文件的唯一表示 方点同样占用磁盘空间 件描述信息从目录项中分离,即应用了索引节点的方法 索引节点可以减少查找文件时的I/0信息量		
<u>索引节点</u>	<u>分类</u>	磁盘索 引节点 内存索	 指存放在磁盘上的索引节点,每个文件有一个唯一的磁盘索引节点 包含内容:文件主标识符;文件类型;文件存取权限;文件物理地址;文件长度;文件链接计数;文件存取时间 指存放在内存中的索引节点,文件打开后,将磁盘索引节点复制到内 		
	书的目录举	<u>引节点</u> 列	存中		
示例图	中国		1 1 1 1 1 1 1 1 1 1		

文件的操作

PALLADINAL				
基本操作	• 创建文件; 写文件; 读文件; 重新定位文件; 删除文件; 截断文件			
文件打开 的过程	 文件打开就是调用open,根据文件名搜索目录,将指明文件的属性(包括该文件在外上的物理地址) 从外存复制到内存打开文件表的一个表目中,并将该表目的编号(也叫索引)返回给用户 【打开文件操作的主要工作就是把指定文件的目录复制到内存指定的区域】 【open操作会把文件的FCB调入内存,而不会把文件内容读到内存,只有进程希望获取文件内容时才会读入文件内容】 			
文件打开和关 闭关联的信息	• 文件指针;文件打开计数;文件磁盘位置;访问权限 文件描述符是打开文件的标识			
	例图	write() sys_write() 文件系统的写方法 用户空间 虚拟文件系统 文件系统		
<u>写入文件</u> <u>的过程</u>	代码	fd = open(name, flag);# 打开文件 write(fd,);# 写数据 read(fd,);# 读数据 close(fd);# 关闭文件		
	解释	• <u>首先用</u> open <u>系统调用打开文件</u> , open <u>的参数中包含文件的路径名和文件名</u> • <u>使用</u> write <u>写数据,其中</u> write <u>使用</u> open <u>所返回的文件描述符,并不使用文件名作为参数</u> 【read同理,结合C语言文件读写代码记忆】 • <u>使用完文件后,要用</u> close <u>系统调用关闭文件,避免资源的泄露</u>		

文件的保护

- 1. 保护的目的:解决对文件的读,写,执行的许可问题
- 2. <u>一个文件的访问常由用户访问权限和文件属性</u>(包括保存在FCB中对文件访问的控制信息)共同设置
- 3. 保护的方式

○ NOT 11/2/2/2/				
		1. 口令	2. 加密保护	
非访问	定义	・ 用户建立一个文件时需要提供口令・ 用户请求访问时必须提供相应口令	• 对文件进行加密,访问时需要密钥	
<u>控制方法</u>	优点	• 时间空间开销不多	• 保密性强,节省了存储空间	
	缺点	• <u>口令直接存在系统内部,不安全</u>	• 编码和译码需要时间	
	 ・ <u>访问控制的目的</u>: 用于控制用户对文件的访问方式 ・ <u>访问控制的对象</u>: 读; 写; 执行;添加;删除;列表清单 ・ <u>访问控制机制必须由系统实现</u> 			
		方法一	方法二	
访问 <u>控制方法</u>	定义	• <u>为每个文件和目录增加一个访问控制列表ACL</u> • <u>该表规定每个用户名及其所允许的空间管理</u>	• <u>采用精简的访问列表</u> • 该列表采用拥有者,组和其他三种 用户类型	
	优点	• 可以使用复杂的访问方法	• <u>只需要三个域即可列出访问表中这</u> 三类用户的访问权限	
	缺点	• 长度无法预计并且可能导致复杂的空间管理		

文件的逻辑结构【用户角度的文件组织形式】

定义	• 即文件中的数据在逻辑层面是如何组织起来的		
无结构文件/ 流式文件	最简单的文件组织形式,是有序相关信息项的集合,以字节为单位对基本信息单元操作不多的文件适合该方式,如源代码文件,目标代码文件		
有结构文件/ 记录式文件	1. 顺序文件	・ <u>串结构: 只能按顺序查找,费时</u> ・ 顺序结构: 可采用折半查找,检索效率高 ・ 顺序查找平均次数 $\frac{N}{2}$; 索引顺序查找平均次数 $N^{\frac{1}{2}}$	
	2. 索引文件	• 提高了存取速度,但索引表增加了存储空间	
	3. 索引顺序文件	• 提同一组中的关键字可以无序,但组与组之间的关键字必须有序 • 先通过索引表查找所在的组,然后再该组中使用顺序查找	

文件的物理结构【文件在外存上的存储组织形式】

✓ITH 3		
<u>定义</u>	 研究文件数据在物理存储设备上是如何分布和组织的 文件在磁带上→连续存放方式 文件在磁盘上→不采用连续存放方式 文件在内存上→随机存放方式 	
	• 文件的存储空间管理就是对磁盘空闲块的管理 • Linux 文件系统就采用了位图的方式来管理空闲空间,不仅用于数据 闲块的管理,因为 inode 也是存储在磁盘的	空闲块的管理,还用于 inode 空
	序号 第一个空 闲	块号 空闲块个数

3 • 为所有空闲空间建立一张表 13 3 • 表内容包括空闲区的第一个块 空闲表法 号和该空闲区的块个数 磁盘块 1 9 10 13 14 15 文件的 存储空 通过链表的方式 将空闲块组织起来 间管理 • 每一个空闲块里有一个指针指 <u>向下一个空闲块</u> 空闲链表法 • 这样能很方便的找到空闲块并 磁盘块 管理起来 空闲块 已分配的块 • 位图是利用二进制的一位来表 • 当值为 0 时,表示对应的盘块空闲 示磁盘中一个盘块的使用情况 • 值为 1 时,表示对应的盘块已分配 位图法 • 磁盘上所有的盘块都有一个二 • 通常可用m*n个位数来构成m行n列的位示图 进制位与之对应 • 且m*n等于磁盘的总块数 成组链法 • 结合空闲表和空闲链表的优点,克服表长的缺点

<mark>文件的物理结构【</mark>文件在外存上的存储组织形式】

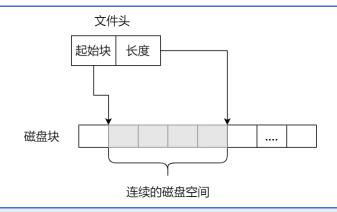
• 文件的存储就是对磁盘非空闲块的管理

方	式	访问磁盘次数	优点	缺点
顺月 分配		需访问磁盘1次	顺序 存取速度快,当文件是 定长时可以根据文件起始地 址及记录长度 进行随机访问	要求连续的存储空间,会产生外部碎 片,不利于文件的动态扩充
链		需访问磁盘n次	无外部碎片,提高了外存空 间的利用率,动态增长比较 方便	智能按照文件的指针链顺序访问,查 找效率低,指针信息存放消耗内存或 磁盘空间
索亞分配		m 级需访问 磁盘 m+1 次	可以 随机访问,易于文件的 增删	索引表增加存储空间的开销,索引表 的查找策略对文件系统效率影响较大

1. 顺序分配【类似数组】

- 文件头需指定起始块的位置和长度
- 要求连续的存储空间
- 可随机访问

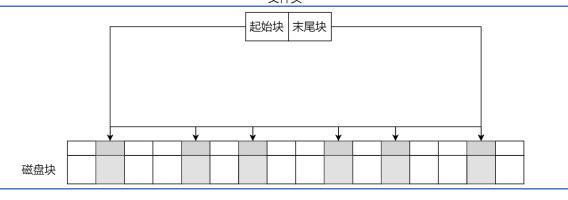
文件的 存储方式



2. 链表分配【类似链表】

- 存放是离散的,不用连续的,于是就可以消除磁盘碎片
- 可大大提高磁盘空间的利用率,同时文件的长度可以动态扩展
- 隐式链接无法直接访问数据块,只能通过指针顺序访问文件
- 显式链接把用于链接文件各数据块的指针,显式地存放在内存的一张链接表中
- 内存中的这样一个表格称为文件分配表FAT (File Allocation Table)
- 不可随机访问

文件头

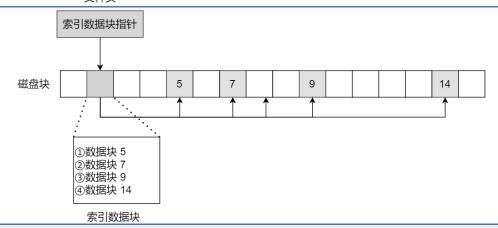


文件的物理结构【文件在外存上的存储组织形式】

3. 索引分配【类似索引表】

- 为每个文件创建一个「索引数据块」
- 里面存放的是指向文件数据块的指针列表
- 像书的目录一样,要找哪个章节的内容,看目录查就可以
- 可随机访问

文件头



4. 链表 + 索引

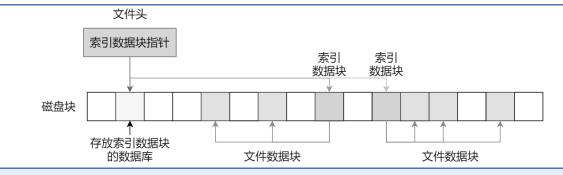
• 索引数据块留出一个存放下一个索引数据块的指针

文件头 索引数据块指针 指向下一个索引数据块 磁盘块

文件的 存储方式

5. 索引+索引【多级索引】

• 通过一个索引块来存放多个索引数据块



多级索引举例→unix系统的inde结构

- 对于小文件使用直接查找的方式可减少索引数据块的开销
- 对于大文件则以多级索引的方式来支持
- 所以大文件在访问数据块时需要大量查询
- 存放文件所需的数据块小于 10 块,则采用直接查找的方式
- 存放文件所需的数据块超过 10 块,则采用一级间接索引方式
- 前面两种方式都不够存放大文件,则采用二级间接索引方式
- 二级间接索引也不够存放大文件,则采用三级间接索引方式

文件系统

基本概念和目标

概念	 文件系统 = 0S中负责管理持久数据的子系统 文件系统 = 与文件管理有关的软件 + 被管理的文件 + 试试文件管理所需的数据结构 文件系统需先挂在到某个目录才可正常使用 文件的基本操作单位就是数据块 		
目标	1. <u>实现对文件的基本操作</u> = 按 <u>名存储和查找文件</u> + <u>组织成合适的结构</u> + <u>文件共享</u> + <u>文件保护</u> 【用户角度】 2. 管理与磁盘的信息交换 + 完成逻辑结构和物理结构的变换【OS角度】 3. 组织文件在磁盘上的存放 + 采取好的文件排放顺序和磁盘调度方法【OS角度】		
	磁盘的文件系统 • 它是直接把数据存储在磁盘中,比如 Ext 2/3/4、XFS 等都是这类文件系统		
<u>分类</u>	内存的文件系统	・ 这类文件系统的数据不是存储在硬盘的,而是占用内存空间・ 我们经常用到的 /proc 和 /sys 文件系统都属于这一类・ 读写这类文件,实际上是读写内核中相关的数据	
	网络的文件系统	• <u>用来访问其他计算机主机数据的文件系统,比如</u> NFS、SMB <u>等等</u>	

文件系统的层次结构

		主要功能和介绍
いの協力	设备驱动程序	• <u>将输入的命令翻译成底层硬件的特定指令</u>
I/O <u>控制</u>	中断处理程序	• <u>利用指令使I0设备与系统交互</u>
基本文 件系统		<u>驱动程序发送通用命令</u> ,以读取和写入磁盘的物理块 区,保存各种文件系统,目录和数据块的缓冲
文件组 织模块	组织文件及其逻辑块和物理块可以将逻辑地址转换为物理地址有空闲空间管理器,以跟踪未分配的块,根据需要提供给文件组织模块	
逻辑文件系统	用于管理元数管理目录结构通过FCB维护文负责文件保护	据信息(包括文件系统的所有结构,不包括文件内容) C件结构

虚拟文件系统VFS

<u>目的</u>	• 为用户程序提供了文件系统操作的统一接口,屏蔽了不同文件系统差异和操作细节
特性	1. <u>能提高系统性能</u> 2. <u>不是一种实际的文件系统</u> 3. <u>只存在于内存中</u> ,不存在与任何外存空间中 4. 在系统启动时建立,在系统关闭时消亡
VFS的数据结构	1. 超级块对象 2. 引节点对象 3. 且录项对象 4. 文件对象

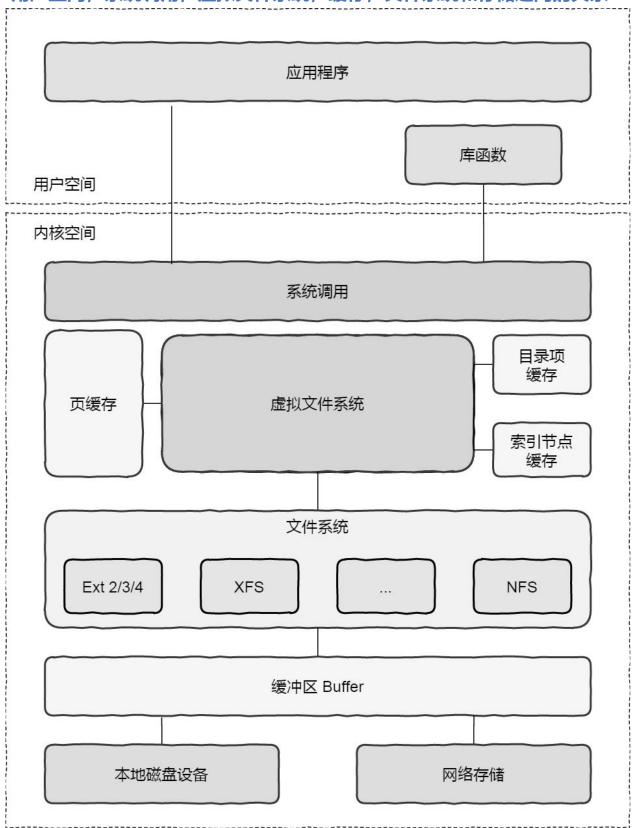
文件系统的布局



分区和安装

- 一个磁盘可划分为多个区,每个分区都可以创建单独的文件系统,每个分区都可包含不同的操作系统
- 文件在使用前必须先安装(即挂载)

用户空间,系统调用,虚拟文件系统,缓存,文件系统和存储之间的关系



目录

常考点

目录结构相关	1. 在树型目录中,为了加快文件检索速度,可设置当前目录,于是文件路径可以从当前目录开始查找 2. 文件系统采用多级目录的目的是解决命令冲突 3. 树形目录结构中,用户对文件的首次访问通常采用文件路径名,之后对文件的访问通常使用文件描述符
<u>目录检索相关</u>	• 在顺序检索法时,只要路径名的一个分量名未找到,就应停止查找
文件共享相关	• 用户进程的打开文件表关于同一个文件不一定相同,例如读写指针位置不一定相同
其他	• 整个文件系统只有一个系统文件打开表(里面的表项都是不重复的),同一文件打开多次只需改变引用计数

目录管理要求

- 1. 实现"按名存取"
- 2. 要提高目录的检索速度
- 3. 需要提供用于控制访问文件的信息
- 4. 允许不同用户对不同文件采用系统的名字

目录的操作

搜索文件 修改目录 创建文件 创建目录 删除目录 移动目录

删除文件:删除一个文件时,会根据FCB回收相应的磁盘空间,将FCB回收,并删除目录中对应的目录项

目录的查询/检索

概念	• <u>目录查询通过在磁盘上反复搜索完成</u> ,需要不断进行I/0操作,开销大 • 可以把当前使用的文件目录复制到内存,从而降低磁盘操作次数,提高系统速度				
实现方法		定义	优点	缺点	
	线性列表 【对应线性查找】	采取线性列表存储文件目录项	实现简单	查找费时	
	哈希表 【对应散列查找】	采取哈希表存储文件目录项	查找迅速,插入删除 简单	需要一些措施来 避免冲突	

目录结构

	<u>単级目录结构</u>	<u>两级目录结构</u>
<u>定义</u>	• 整个文件系统只建立一张目录表 • 每个文件占一个目录项	 文件目录分为主文件目录MDF和用户文件目录 UFD MDF记录用户名UFD所在的存储位置 UFD记录用户文件的FCB信息
<u>优点</u>		解决了多用户之间的文件重名问题文件系统可以在目录上实现访问限制
缺点	・ <u>查找速度慢</u>・ 文件不允许重名・ 不便于文件共享・ 不适合多用户的OS	• <u>缺乏灵活性</u> ,不能对文件分类
结构图	FCB1 FCB2 FCB3 FCBn 文件1 文件2 文件3 文件n 图 4.11 单级目录结构	User ₁ User ₂ User ₃ User ₄ 主文件目录 CAT BO A TEST A DATA X DATA A 用户文件目录 文件 图 4.12 两级目录结构
	树形目录结构	无环图目录结构
定义	・使用绝对路径,相对路径,当前路径的结构・不同的用户的文件,文件名可相同可不同・大多0S采用这种目录结构	• <u>在树形目录结构上</u> • 加入有向边,组成一个有向无环图
<u>优点</u>	• 可以很方便的对文件进行分类 • 能够有效地进行文件的管理和保护	• <u>实现了文件共享</u>
缺点	• <u>利于文件共享</u> • <u>查找文件增加了磁盘访问次数</u> ,会影响查 询速度	• 使系统的管理变得更加复杂
结构图	dev home wsr ws	root root spell count list p list R 4.14 无环图目录结构

文件共享

概念	• 文件共享使多个用户共享同一个文件,系统只需保留该文件的一个副本			
文件共享方式		基于索引节点的关系方式【硬链接】	基于符号链实现文件共享【软链接】	
	定义	• <u>硬链接就是多个指针指向一个索引节点</u> • 文件的物理地址和其他文件属性信息放 在索引节点中	• 软链接相当于重新创建一个文件 • 新文件只包含被链接文件的路径名	
	特点	・ 硬链接不可用于跨文件系统・ <u>硬链接查找速度比软链接快</u>	• 软链接可以跨文件系统	
	新增文件时	• 建立硬链接时,引用计数值加1	• 符号链接, 计数值直接复制	
	删除文件时	• 删除文件时,计数值减1 • 若得到的值不为0,则不能删除此文件 • 即只要还有一个指针在,索引节点就不 会被删除	删除操作对符号链接不可见以后再通过符号链接访问时,若发现文件不存在,直接删除符号链接	
	<u>示例</u>	Wang用户文件目录 Test r 常引结点 Count=2 文件物理地址 Test r Each Test r And Test r Test r Each And Test r And Test r Test r	快捷方式属于文件共享中的软链接	