



华中农业大学
HUAZHONG AGRICULTURAL UNIVERSITY

勤讀力耕
立己達人

操作系统原理

 **Principle of Operating System**

主讲：倪福川

E-mail: fcni_cn@mail.hzau.edu.cn

TEL:18602714039

信息学院 大数据科学系



第七章 文件管理

- 第一节 文件和文件系统
- 第二节 文件的逻辑结构
- 第三节 文件目录
- 第四节 文件共享



7.1 文件和文件系统

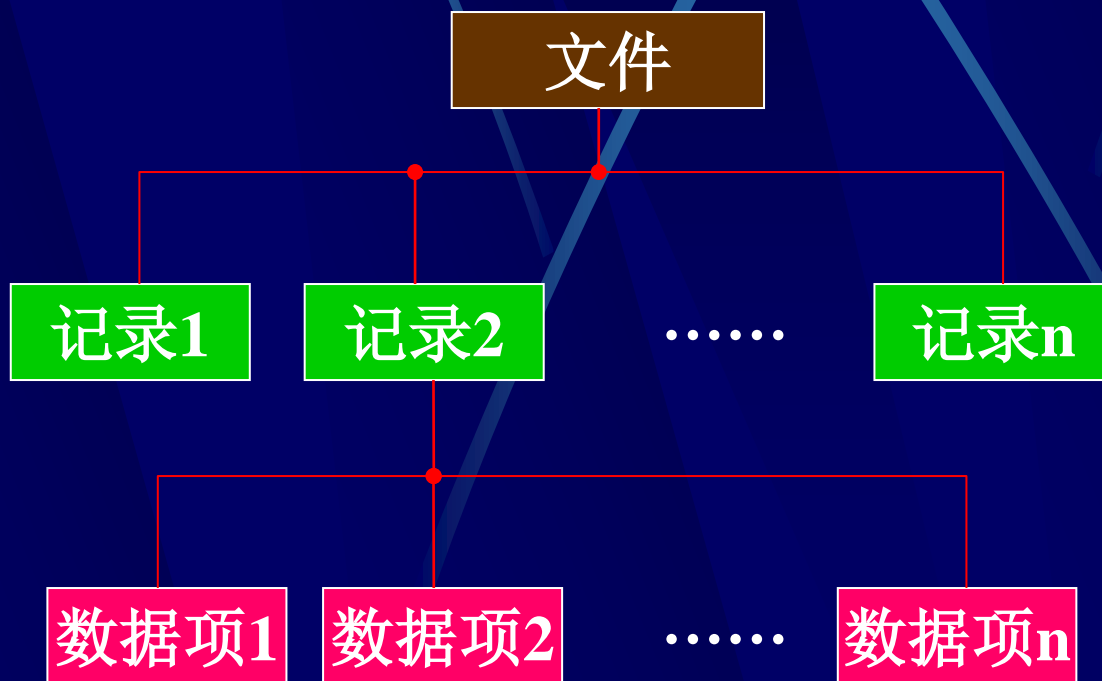
OS通过文件系统来组织和管理计算机中存储的大量数据和程序。

- 7.1.1 文件、记录和数据项
- 7.1.2 文件类型和文件系统模型
- 7.1.3 文件操作



7.1.1、文件、记录和数据项

- 基于文件系统的概念，可以把数据组成分为数据项、记录 and 文件三级





例1：学生成绩单

文件

数据项

学号	姓名	语文	数学	英语	物理	化学
001	张三	90	85	79	92	78
002	李四	86	91	86	90	82
003	王五	88	87	84	89	75
004	赵六	78	92	95	87	91
005	周七	85	88	96	90	90

记录



●1.数据项

- 基本数据项：描述一个对象的**某种**属性的字符集
- 组合数据项：由**若干个**基本数据项组成

●2.记录

- 记录是一组相关**数据项**的集合，用于描述一个对象的某些属性。
- 关键字：能够**唯一**标识一个记录的数据项





- **3.文件**是指由创建者所定义的、具有文件名的一组相关数据元素的集合;
- ☆● **文件的属性:**
文件类型、文件长度、文件的物理位置、文件的建立时间等。



7.1.2、文件类型和文件系统模型

●1、文件的类型

1) 按文件的性质和用途分:

- 系统文件：由系统软件构成的文件，只允许调用执行，不允许用户读和修改。
- 用户文件：只允许文件的授权者使用。
- 库文件：允许用户调用不允许修改。



2) 按文件中数据的形式分:

源文件、目标文件、可执行文件

3) 按存取控制属性分:

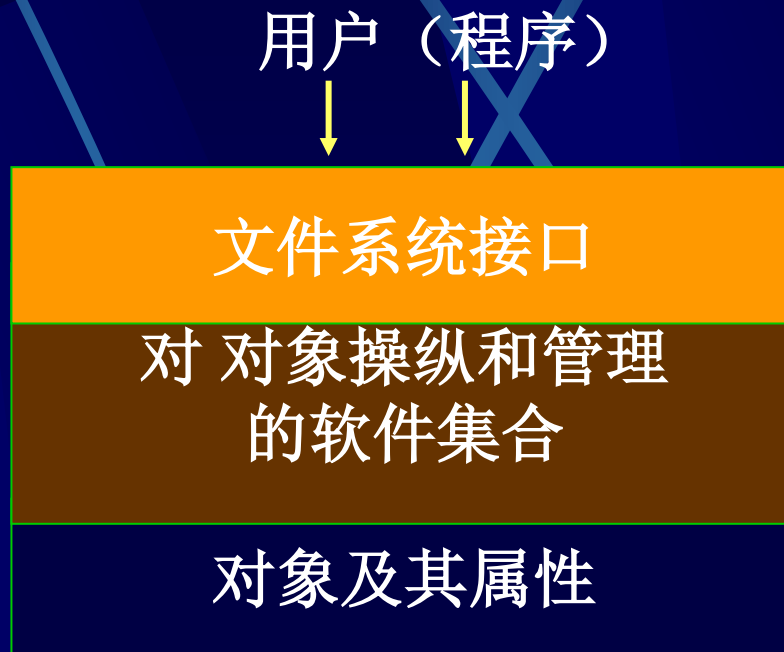
- 只执行文件、只读文件、读写文件

4) 按组织形式和处理方式分:

- 普通文件: ASCII码或二进制码组成的字符文件
- 目录文件: 由文件目录组成
- 特殊文件: 系统中的各类I/O设备



2、文件系统模型



1) 对象及其属性

- 文件：文件管理的直接对象
- 目录：方便用户对文件的存取和检索
- 磁盘（磁带）存储空间



例：MS-DOS的目录结构

文件名	扩展名	属性	备用	时间	日期	首盘块号	盘块数
f1	exe	只读
f2
f3							
f4							



文件系统接口

对对象操纵和管理的 软件集合

对象及其属性

- 2) 对对象操纵和管理的软件集合（核心）
 - 功能：对文件存储空间、文件目录的管理、地址转换机制、文件读写、文件的共享与保护。
- 3) 文件系统的接口
 - 命令接口：用户与文件系统的接口
 - 程序接口：用户程序与文件系统的接口



7.1.3、文件操作

- 用户通过文件系统提供的**系统调用**实施对文件的操作
- 1、最基本的文件操作
 - 1) **创建文件**: 分配外存空间→建立目录项
 - 2) **删除文件**: 删除目录项→回收外存空间
 - 3) **读文件**: 文件名、内存目标地址、目录项、读指针
 - 4) **写文件**: 文件名、内存中源地址、目录项、写指针



●2、文件的打开与关闭

- 打开**：系统将指名文件的属性（包括文件在外存的物理位置）从外存拷贝到内存**打开文件表**的一个表目中，将表目编号返回用户
- 关闭**：将文件从打开文件表的表目上删除，释放表目空间

●3、其它操作

- 对文件属性的操作**：改变文件名、文件主、访问权
- 对文件目录的操作**：创建、删除目录等



7.2 文件的逻辑结构

文件逻辑结构是从用户角度观察到的文件组织形式
文件物理结构是文件在外存上的存储组织形式

- 文件逻辑结构的类型
- 顺序文件
- 索引文件
- 索引顺序文件



对文件逻辑结构提出的基本要求：

- 提高检索速度
- 便于修改
- 减少文件占用的存储空间



7.2.1、文件逻辑结构的类型

● 1、有结构文件（记录式文件）

- 1) **定义**：由一个以上的记录构成的文件
- 2) **基本分类**：定长记录、变长记录
- 3) **文件的组织**：
 - 顺序文件：一系列记录按某种顺序排列形成
 - 索引文件：记录为变长，**每个**记录一个索引表项
 - 索引顺序文件：**每组**记录的第一个记录设一表项



●2、无结构文件（流式文件）

- 定义：**由字符流构成的文件
- 大量的源程序、可执行文件、库函数等
- 文件长度以**字节**为单位
- 对流式文件的访问采用**读写指针**指出下一个要访问的字符
- UNIX**系统中所有文件都被看作是流式文件



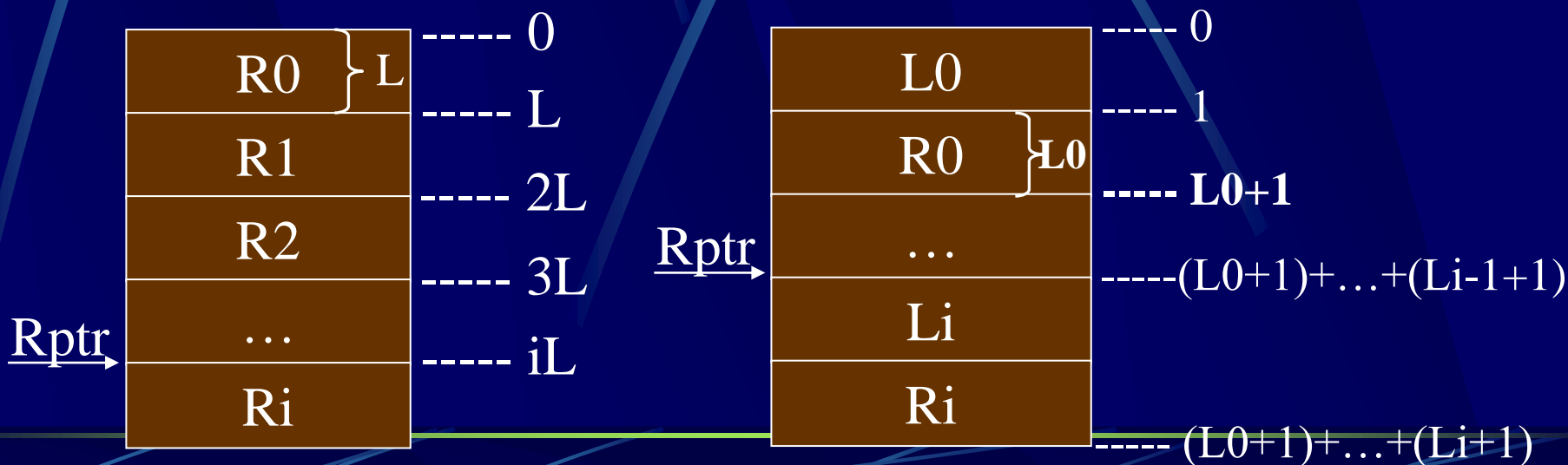
7.2.2、顺序文件

1、逻辑记录的排序

- 串结构 (以时间排序)、顺序结构 (按关键字排序)

2、对顺序文件的读/写操作

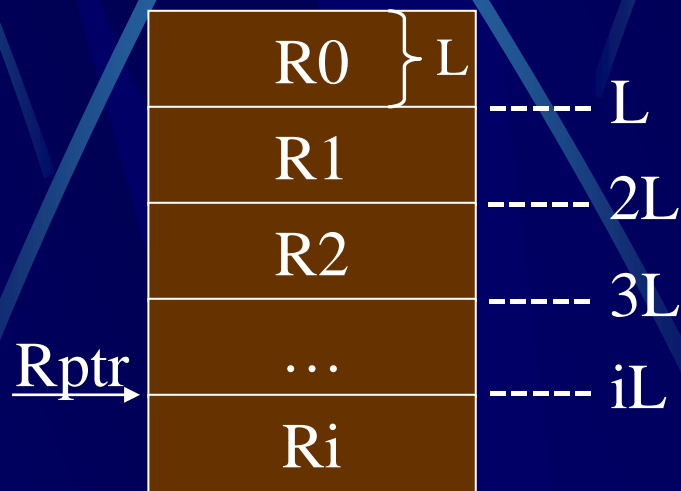
- 1) 定长记录: $Rptr := Rptr + L$
- 2) 变长记录: $Rptr := Rptr + L_i$





●3、顺序文件的优缺点：

- 1) 优点：适于批量存取、能用于磁带存储
- 2) 缺点：查找/修改/增/删单个记录效率低，系统开销大



顺序文件



7.2.3、索引文件

- 1、利用定长记录的顺序文件访问变长记录的文件
- 2、索引表本身是一个定长记录的顺序文件
 - 索引号（记录键或关键字）
 - 长度
 - 指针

检索时，文件用增加顺序记录的关键字和查找算法，检索索引表，访问主文件的记录。

索引号	长度	指针
0	m0	xx
1	m1	xx
2	m2	xx
...
i	mi	xx
...
...
...

R0
R1
R2
...
Ri
...
...
...

索引表

逻辑文件



7.2.4、索引顺序文件

- **索引顺序文件**是顺序文件和索引文件的结合，是最常见的一种逻辑文件形式。
- **原理：**
 - 1) 顺序文件中的所有记录分为**若干个组**；
 - 2) 为顺序文件建立一张索引表，在索引表中为每组中的第一个记录建立一个索引项；



索引键	逻辑地址
AAA	
BAAA	
CAA	
.....
.....
.....
.....
.....

索引表

主键	其它属性
AAA	
AABD	
AAC	
.....
BAAA	
.....
CAA	
.....

逻辑文件

$$\left(\frac{n}{m} + m\right) 2$$

P₂₃₁

平均查找 \sqrt{N}

检索时，利用用户程序提供的关键字和查找算法，
检索索引表，...，利用顺序查找法查找主文件



7.2.5、直接文件和哈希文件

● 1、直接文件

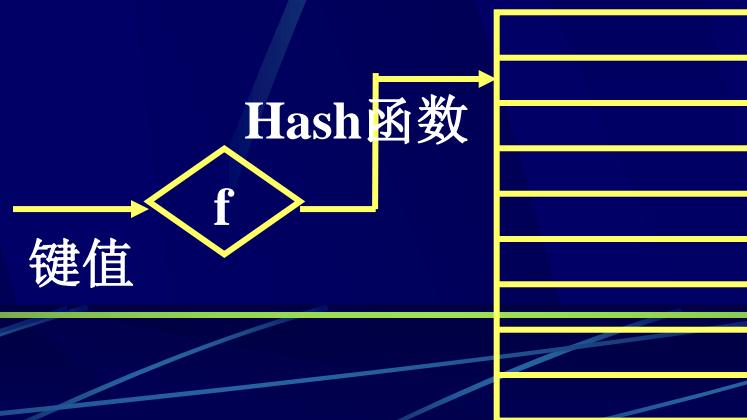
- 前述文件结构对记录进行存取时，都需利用给定的记录键值（关键字），对线性表或链表进行建设，以找到指定记录的物理地址。
- **直接文件**：根据给定的记录键值，直接获得物理地址。即记录键值本身决定了记录的物理地址
- **键值转换**——由记录键值到记录物理地址的转换



● 2、哈希文件

- 是目前应用最广泛的一种直接文件。利用hash函数，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，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。
- 例如，Hash函数 $A=H(K)$

K:记录键值 A: 该记录在目录表中对应表目的位置





7.3 文件目录

- 目录管理的要求
- 文件控制块和索引结点
- 目录结构
- 目录查询技术



目录管理的要求

- **目录**：用于标识系统中文件及其物理地址的一种数据结构，供检索使用。
- **目录管理的要求**：
 - 实现“**按名存取**”——最基本的功能
 - 提高对目录的检索速度
 - 文件共享
 - 允许文件重名



7.3.1、文件控制块和索引结点

● 1、文件控制块（FCB）

- **定义：**描述和控制文件的数据结构
- **FCB的有序集合称为文件目录（或目录文件）**
- **FCB包含的信息项**
 - **基本信息：**文件名/物理位置/逻辑结构/物理结构
 - **存取控制信息：**不同用户存取权限不同
 - **使用信息：**建立/修改的日期、时间等

MS-DOS
的FCB

文件名	扩展名	属性	备用	时间	日期	第一块号	盘块数



●2、索引结点

●1) 索引结点的引入:

- 文件很多时，目录文件占用大量盘块。查找目录时要多次启动磁盘，顺序读取存放目录文件的盘块。
- 实际上，检索目录文件时，只需要利用文件名进行查找，所以可以给文件目录瘦身。
- UNIX系统中，把文件名和文件描述信息分开，由文件描述信息单独构成索引结点（简称i结点）
- FCB改变为：文件名+指向i结点的指针



文件名	索引结点编号
文件名1	
文件名2	
...	...

0 13 14 15

UNIX的文件目录



●2) 磁盘索引结点：（每个文件唯一）

文件主标识符：标识文件拥有者

文件类型

文件存取权限

文件物理地址：每个索引结点有13个地址项，直接或间接给出盘块号

文件长度

文件连接计数：

所有指向该文件的文件名的指针计数

文件存取时间：文件最近被存取修改的时间和索引结点被修改的时间。

索引表
定长的叫直接索引



●3) 内存索引结点:

存放在内存中的索引结点。

当文件被打开时，要将磁盘的索引结点拷贝到内存的索引结点中。

增加了以下内容:

索引结点编号、状态、访问计数、文件所属文件系统的逻辑设备号、链接指针



7.3.2、目录结构

● 1、单级目录结构

- 在整个文件系统中建立一张目录表，每个文件占一个目录项。
- **目录项包括：**文件名、扩展名、长度、类型、物理地址及其它文件属性。

文件名	物理地址	文件说明	状态位
文件名1	XXXX	XXXX	XXXX
文件名2	XXXX	XXXX	XXXX
.....

表明目录项
是否空闲



• 单级目录结构

- 新建文件时，要检索所有目录项，保证新文件名唯一。建立新表项，置状态位为1。
- 删除文件时，找到目录项，回收空间，清除目录项
- 优点：简单、能实现按名存取

文件名	物理地址	文件说明	状态位
文件名1	XXXX	XXXX	XXXX
文件名2	XXXX	XXXX	XXXX
.....

表明目录项
是否空闲

不便于文件共享，只能适用于单用户环境。



2、两级目录结构

- 为每个用户建立一个单独的用户文件目录，由用户的所有FCB组成。
- 系统中再建立一个主文件目录。每个用户的目录文件...

MFD

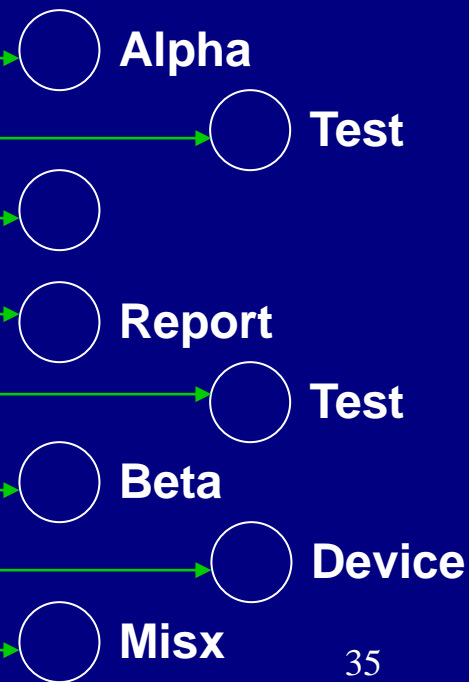
用户名	指向子目录指针
Zhang	xxxx
Wang	xxxx
Sun	xxxx
.....

UFD

Alpha	
Test	

Report	
Test	

Beta	
Device	
misx	





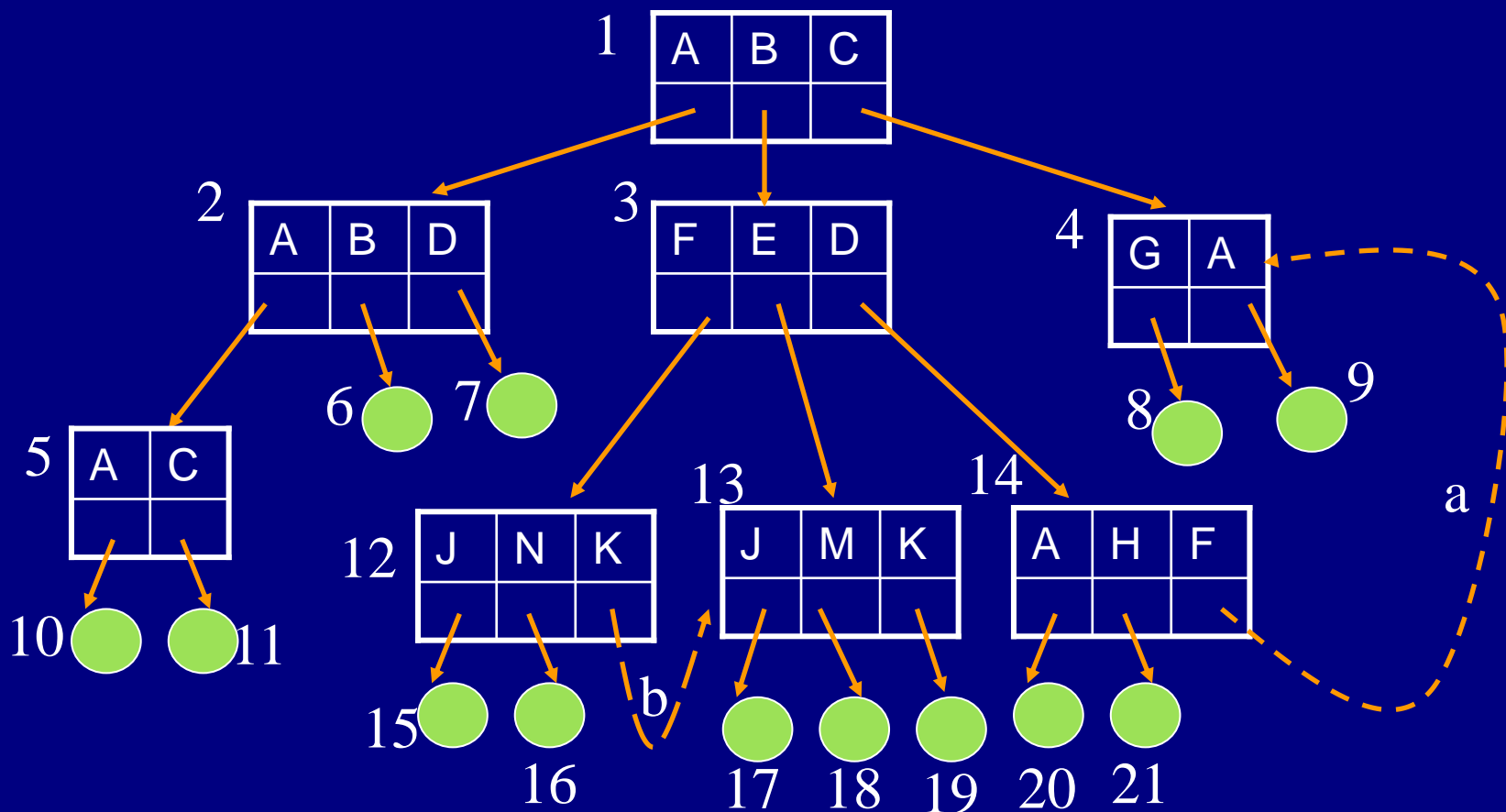
• 两级目录结构

- 用户可请求系统为自己**建立UFD**，也可请求系统将其**撤消**。
- **新建**用户文件时，OS检查用户UFD—>文件名，建立新表项。
- **删除**用户文件时，OS查找用户UFD—>文件目录项，回收存储空间，删除目录项。
- **优点：**
提高检索目录的速度
不同的用户目录中文件可以同名



• 3、多级目录结构（树型目录结构）

• 1) 目录结构:





• 多级目录结构

● 2) 路径名:

从根目录到任何数据文件，都有一条**唯一**的通路，把全部目录文件名和数据文件名，依次用“/”链接起来，构成该数据文件的路径名（绝对路径）。如：

B/E/J

• 3) 当前目录:

进程在一定时间内所访的文件仅限于某个文件目录之下，该文件目录可设置为当前目录。

把从当前目录开始直到数据文件为止构成的路径名叫做**相对路径**。例如：用户B的当前目录是E，则可使用相对路径名“J”来访问自己的J文件。



• 4) 增加和删除目录

● 增加目录:

用户可为自己建立UFD，并可再创建子目录。创建文件时，先查看自己的UFD及其子目录，无重名文件则在UFD或某子目录中增加一个新目录项。

● 删除目录:

若为空目录，可直接删除目录项。

若非空目录，则：

a、不删除非空目录。采用递归方式删除。

b、可删除非空目录。所有文件/子目录同时删除。

MS-DOS



7.3.3、目录查询技术

- 目录查询技术一

- 线性检索法（顺序检索）：

- a. 单级目录中，利用用户提供的文件名，利用顺序查找法，从文件目录中找到指名文件的目录项。
- b. 在树型目录中，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是`/usr/ast/mbox`，则查找过程如下：

根目录文件

1	.
1	..
4	Bin
7	Dev
14	Lib
9	Etc
6	Usr
8	tmp

132

/usr的目录文件

6	.
1	..
19	dick
30	Erik
51	Jim
26	Ast
45	bal

496

/usr/ast的目录文件

26	.
6	..
64	grant
92	book
60	mbox
81	minik
17	src



- 目录查询技术二

- **Hash 方法:**

建立一张Hash索引文件目录后，便可利用Hash方法进行查询。即系统利用用户提供的文件名并将它变换为文件目录的索引值，再利用该索引值到目录中去查找。

- **Hash 冲突**—— n 个不同的文件名有可能转换为相同的Hash值



- 一种处理冲突的有效规则:
- (1) 利用Hash法索引查找目录时, 目录项空, 表示无文件
- (2) 如果目录项中的文件名与指定文件名匹配, 表示找到文件, 并找出文件的物理地址。
- (3) 如果目录项中的文件名与指定文件名不匹配, 即“冲突”。需将Hash值在加上一个常数, 形成新的索引值, 再返回第一步重新开始查找。



7.4 文件共享

- 基于索引结点的共享方式
- 利用符号链实现文件共享



7.4.1、基于索引结点的共享方式

C目录文件

物理地址

文件名1	盘块号	盘块数
文件名2	盘块号	盘块数
filec	盘块号	盘块数
.....

B目录文件

文件名1	盘块号	盘块数
文件名2	盘块号	盘块数
		
filec	盘块号	盘块数

COPY



● 基于索引结点的共享方式

- 文件目录中只设置文件名及指向相应索引结点的指针；
- 文件的物理地址及其它的文件属性等信息只存放在索引结点中；

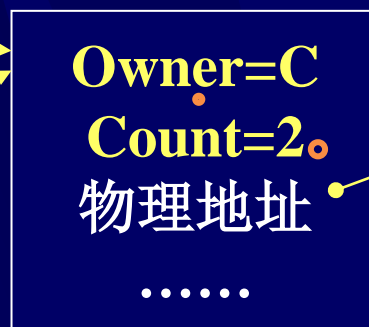
C用户文件目录

CmnFile1	●

B用户文件目录

CmnFile1	●

索引结点



● CmnFile1

链接计数

缺点：用户C不再需要此文件时不能执行删除



7.4.2、利用符号链实现文件共享

● 符号链与符号链接

- 由系统创建一个Link类型新文件，新文件中包含被链接（共享）文件的路径名。
- 要访问共享文件，则必须先访问Link新文件，此时将被OS截获，OS根据新文件中的路径名去读共享文件，这种方式称为符号链接。
- 只有文件主（文件创建者）才拥有指向文件索引结点的指针。因此，文件主可随时删除此文件



C的目录

...	
filec	
...	

B的目录

...	
fileb	
...	

Owner=C

类型：普通

文件物理地址

...

文件filec

Owner=B

类型：LINK

文件物理地址

...

文件fileb

符号链接

符号链

● 优缺点：

- 能够链接任何地方的文件
- 按路径查找的访问开销大，需要冗余的索引结点
- 允许使用多个不同的名字访问共享文件，因此遍历文件系统时会重复访问此共享文件。