计算机操作系统

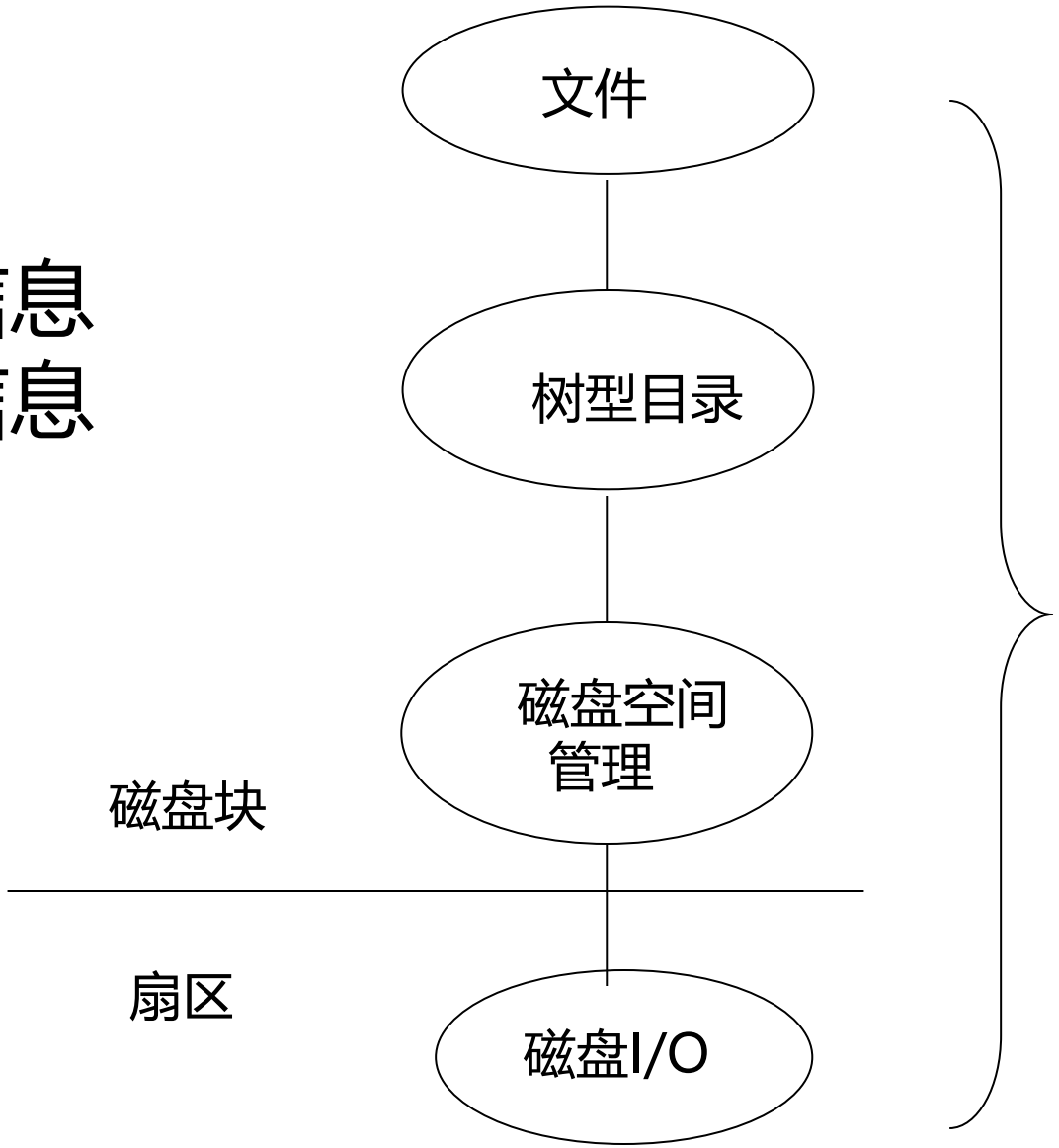
Operating Systems

李 琳

第八章 磁盘存储器的管理

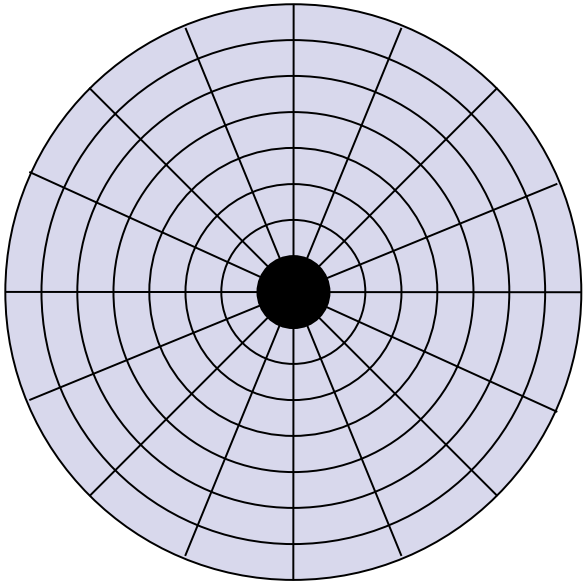
目的

（1）存储信息
（2）找到信息



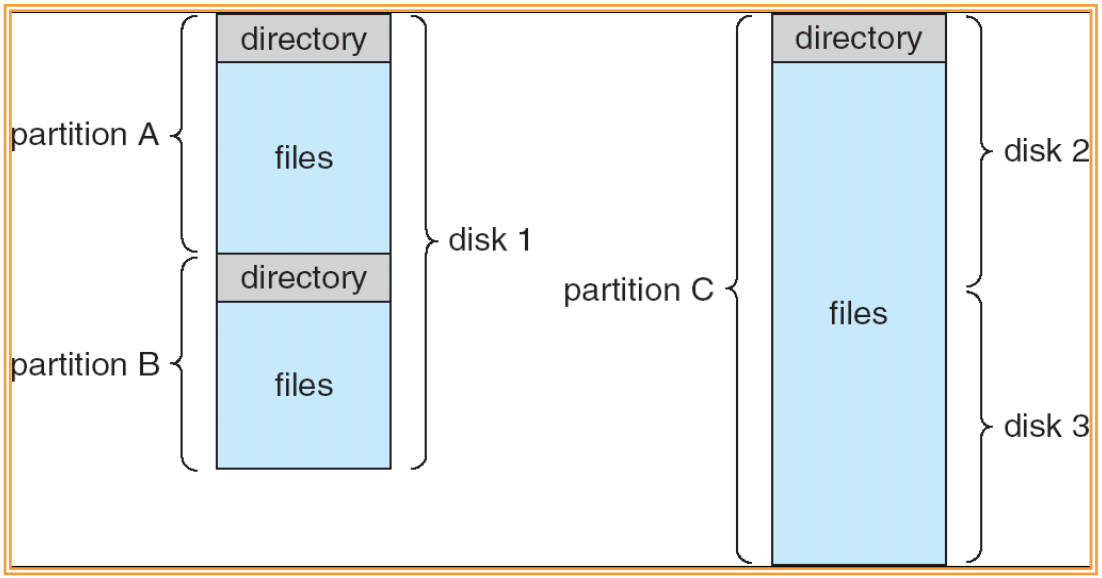
8.1 外存的组织方式

- 文件的物理结构
  - ✓物理单位——扇区
  - ✓逻辑单位——磁**盘块**

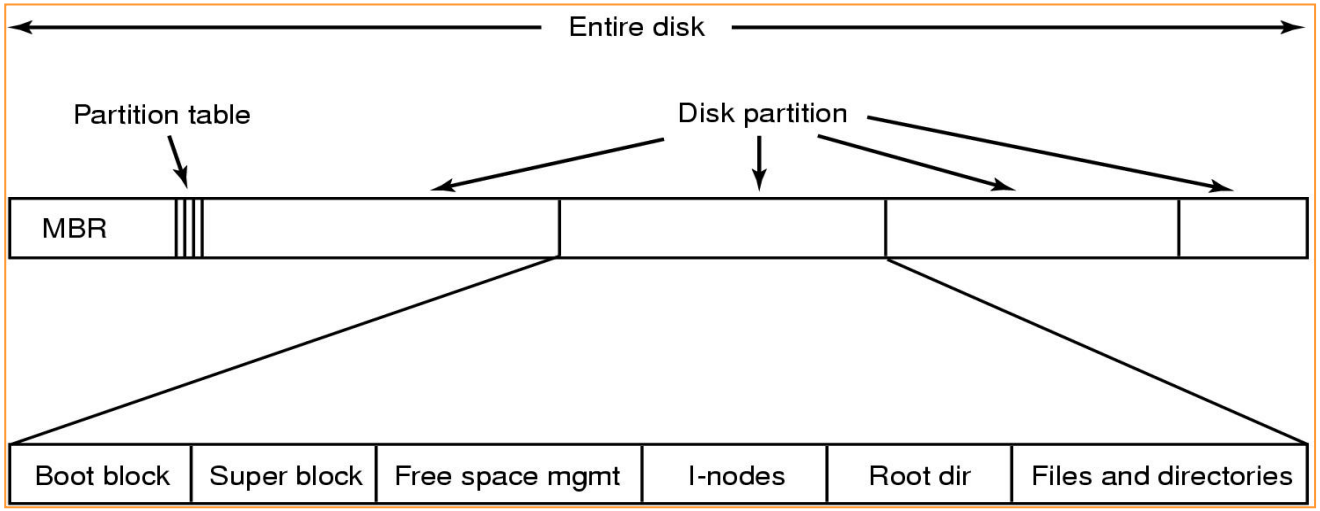


- 磁盘格式化的主要工作：(1) 设置磁盘块大小；(2)为磁盘块编号；

8.1 外存的组织方式



磁盘分区



分区结构

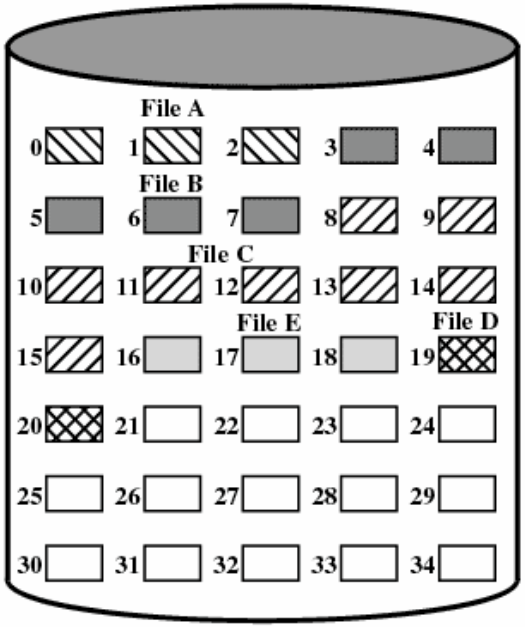
8.1 外存的组织方式

8.1.1 连续分配

- 基本思想
  - ✓要求为每个文件分配一组相邻接的磁盘块，且文件的逻辑记录的顺序与所存储磁盘块的块号顺序一致。
  - ✓所形成的文件结构称顺序文件结构，物理文件称为顺序文件。

- 特点
  - ✓顺序访问速度快
  - ✓要求连续的存储空间
  - ✓文件长度变化很困难

FCB中的外存位置记录第一个盘块号！



File Allocation Table

File Name	Start Block	Length
File A	0	3
File B	3	5
File C	8	8
File D	19	2
File E	16	3

8.1 外存的组织方式

8.1.1 链接分配

- 基本思想
  - ✓采用离散分配思想，为每个文件分配一组不相邻接的磁盘块，且文件的逻辑记录的顺序与所存储磁盘块的块号顺序可不一致。

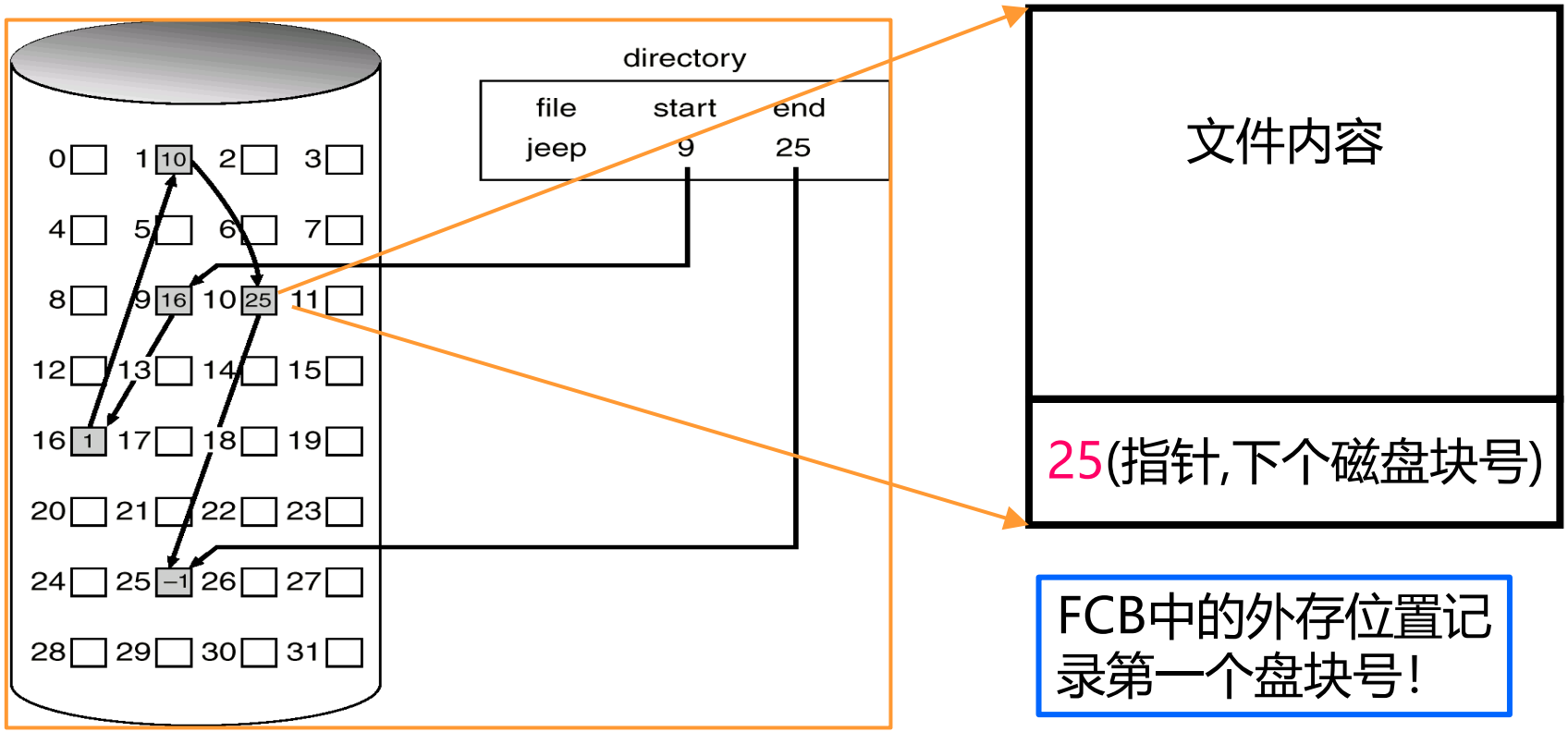
核心问题：如何设计数据结构，记录文件分配到的磁盘块信息？

8.1 外存的组织方式

8.1.2 链接分配——隐式链接

- 隐式链接

  采用单链表结构，各储磁盘块存储指向下个磁盘块的块号。所形成的文件结构称链接文件结构，物理文件称为链接文件。



8.1 外存的组织方式

8.1.2 链接分配——隐式链接

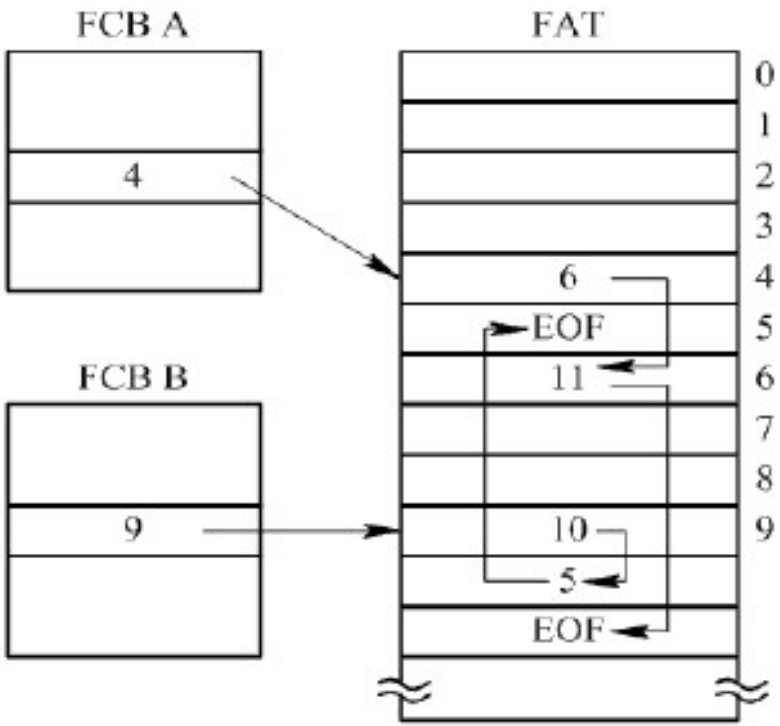
- 隐式链接的特点
  - ✓不存在外部碎片问题
  - ✓有利于文件动态变化
  - ✓读写信息时需要依次访问前序磁盘，存取速度慢
  - ✓不适于随机存取
  - ✓可靠性容错性差，如指针出错

8.1 外存的组织方式

8.1.3 链接分配——显式链接

- 显式链接

设置文件分配表(File Allocation Table, FAT),集中存储所有的磁盘块号信息。



示例：

文件A,占据磁盘块为4,6,11。
文件B,占据磁盘块为9,10,5。

核心：单链表的集中存储。

FCB中的外存位置记录第一个盘块号！

8.1 外存的组织方式

8.1.2 链接分配——显式链接

- 显式链接的特点
  - ✓不存在外部碎片问题
  - ✓有利于文件动态变化
  - ✓读写信息之前需要访问FAT表（内存），存取速度稍慢
  - ✓可以进行随机存取
  - ✓可靠性容错性比隐式链接好

集中存储，因而可以备份FAT表

8.1 外存的组织方式

8.1.3 链接分配——显式链接

FAT类型	FAT占用空间	磁盘块(簇)尺寸	管理磁盘空间
FAT12 12位磁盘块号： [0，2^{12}-1]	1.5*4K = 6K	0.5K 1K	2M 4M
FAT16 16位磁盘块 [0，2^{16}-1]			
FAT32 32位磁盘块号： [0，2^{32}-1]	4*4G = 16G 4*4M = 16M 4*4M = 16M 4*16M= 64M	1K 1K 4K 4K	4T 4G 16G 64G

磁盘大小、盘块大小、FAT表大小三者关系:

盘块个数=磁盘大小/盘块大小

$2^{盘块号字节数*8}$ 》= 盘块个数

FAT表大小=盘块个数*盘块号字节数

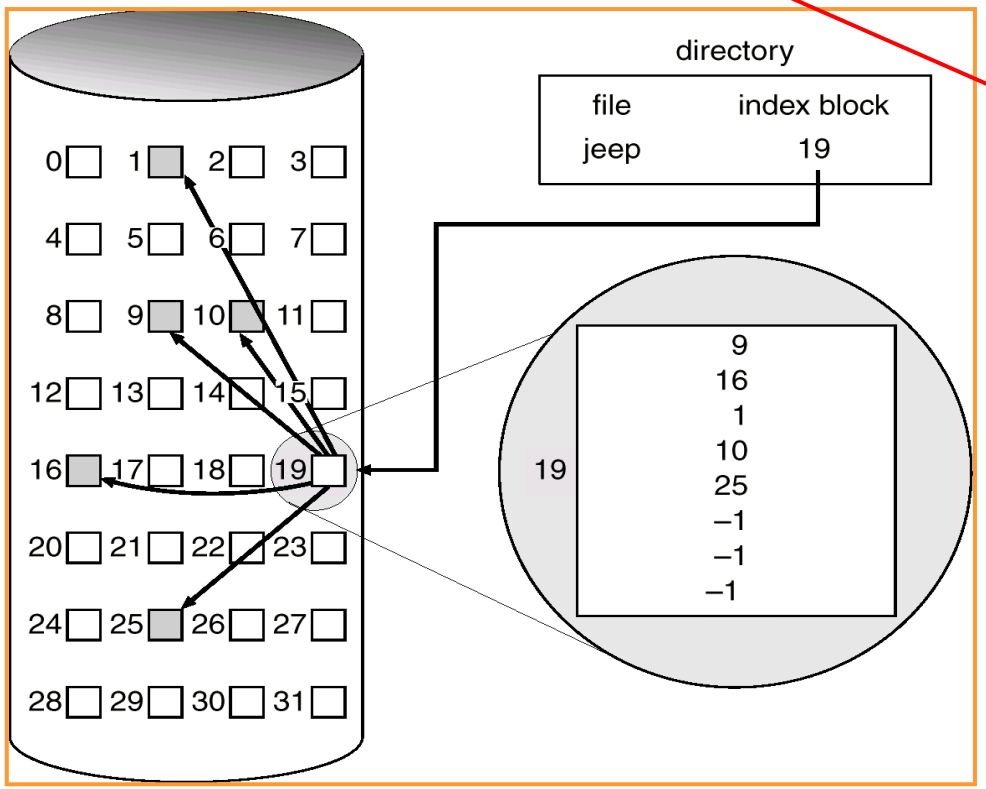
8.1 外存的组织方式

8.1.4 索引分配

- 基本思想

  为每个文件设置索引表，存储其分配到的磁盘块号

  所形成的文件结构称索引文件结构，物理文件称为索引文件



索引块

为了便于保存索引表，将其存储在一个盘块中，形成索引块，区别于普通的文件块。

FCB中的外存位置记录索引块号！

8.1 外存的组织方式

8.1.4 索引分配

- 单级索引

假设磁盘块尺寸为4KB，磁盘块号为32bits。

则每个磁盘块（索引块）可以存储磁盘块号的数量为：

4K/4 = 1K (个);

则单级索引支持的文件尺寸为：

1K * 4K = 4 M。

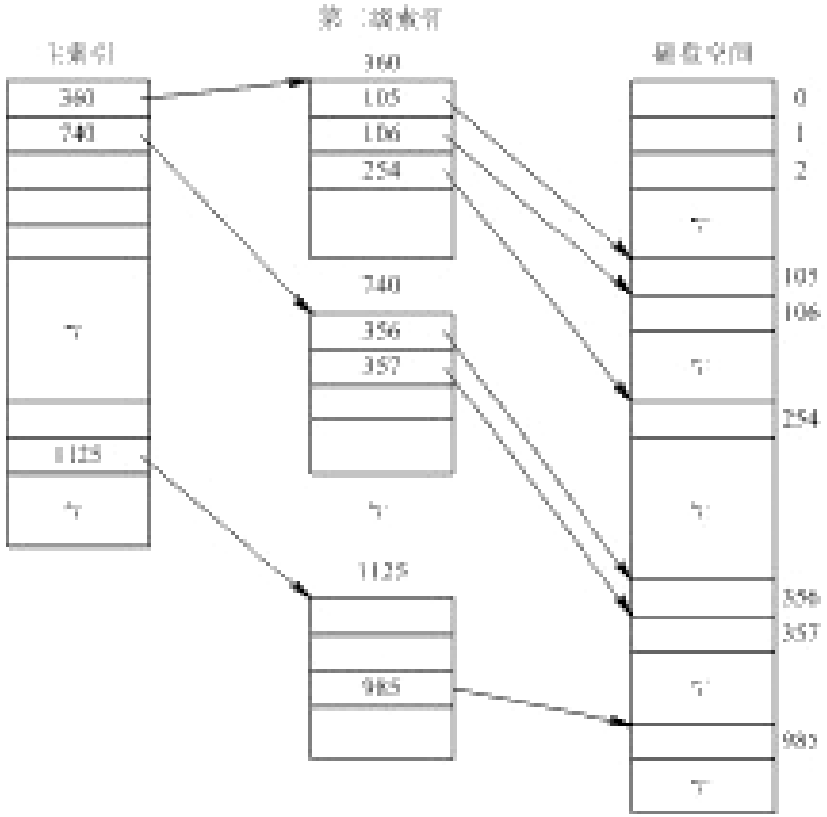
注：小于4M文件的索引表只需要占1磁盘块，但大于4M的文件的索引表无法由1个磁盘块存储，因此这样的系统不支持大于4M的文件。

8.1 外存的组织方式

8.1.4 索引分配

- 二级和多级索引

对于索引表的存储，也采用索引表来记录其所占用的磁盘块号，从而形成二级索引。（大索引表离散化）



假设磁盘块尺寸为4KB，磁盘块号为32bits。

则每个磁盘块可以存储磁盘块号：

4K/4 = 1K （个）；

而二级索引支持的文件尺寸为：

1K * 1K * 4K = 4G。

索引表占用空间：

4K + 1K*4K

1+1K个索引块

8.1 外存的组织方式

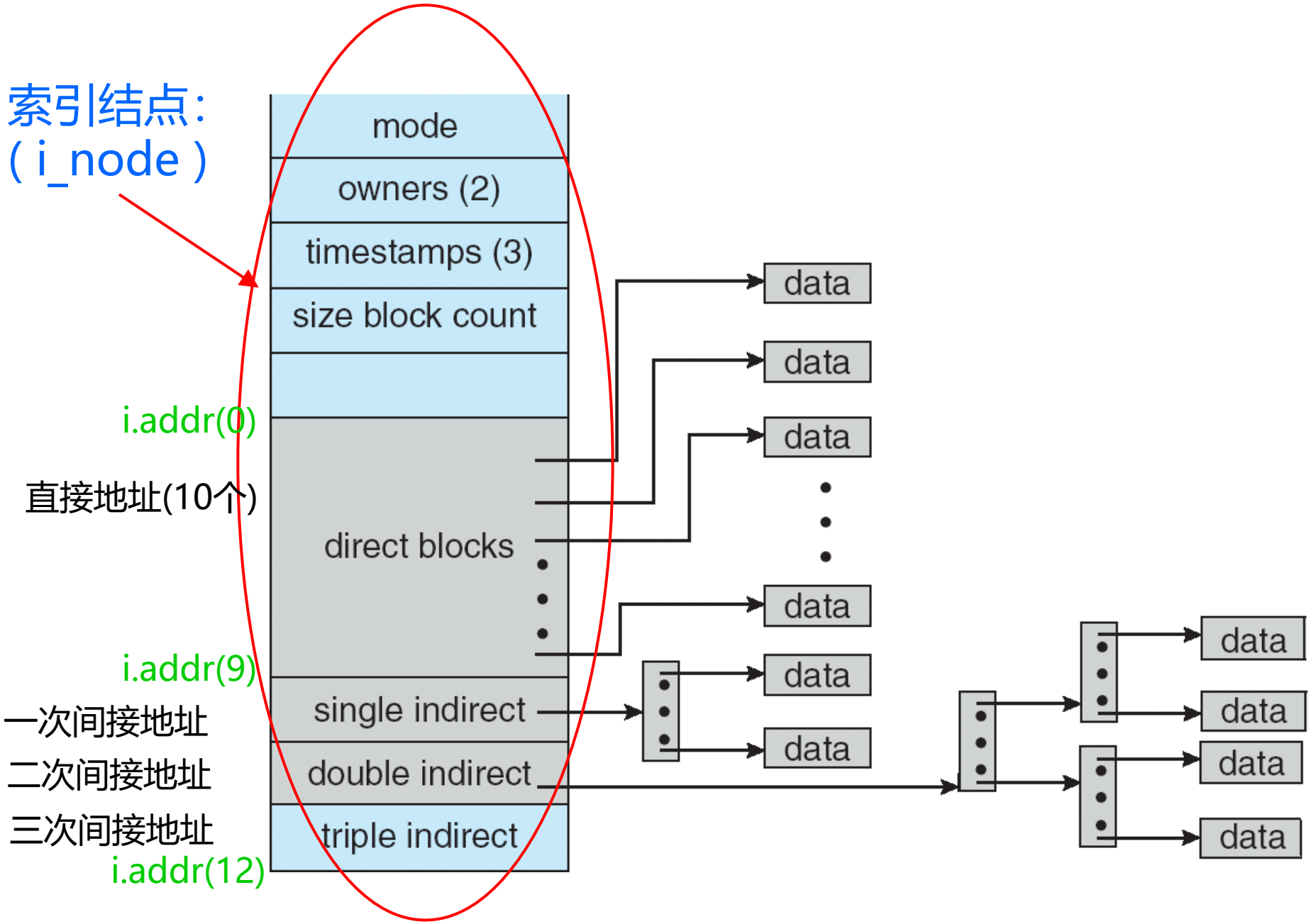
8.1.4 索引分配

- 混合索引分配方式（UNIX System V）

  混合使用直接磁盘块号、各级索引表，从而可以既支持小文件的存储，也可以支持大文件的存储。

  达到降低索引表存储空间和同时支持大中小文件的目的。

- 支持的文件尺寸
  - ✓假设磁盘块尺寸为4KB，磁盘块号为32bits。
  - ✓直接地址, 支持文件尺寸：10*4K = 40K ;
  - ✓一次间接地址, 支持文件尺寸：1K*4K = 4M ;总计：40K + 4M ;
  - ✓二次间接地址, 支持文件尺寸：1K*1K*4K = 4G ;总计：40K + 4M + 4G;
  - ✓三次间接地址, 支持文件尺寸：1K*1K*1K*4K = 4T ;总计：40K + 4M + 4G + 4T;


索引结点:
(i_node)
mode
owners (2)
timestamps (3)
size block count
i.addr(0)
直接地址(10个)
direct blocks
i.addr(9)
一次间接地址
single indirect
二次间接地址
double indirect
三次间接地址
triple indirect
i.addr(12)
data
data
data
data
data
data
data
data
data
data

8.1 外存的组织方式

8.1.4 索引分配

- 索引分配的特点
  - ✓不存在外部碎片问题
  - ✓有利于文件动态变化
  - ✓读写信息前需要访问各个索引块，访问速度慢
  - ✓顺序存取与随机存取并存
  - ✓文件索引互不影响，安全性容错性提高
  - ✓索引表本身带来的系统开销（时间、空间）

8.1 外存的组织方式

	连续分配方法	隐式链接方法	显式链接方法	索引分配方法
动态增删	困难	简单	简单	简单
检索速度	很快	较慢	较快	一般
存取速度	很快	一般	一般	一般
额外空间	不需要	需要	需要	需要
可靠性	很好	较差	较好	一般

试一试

1、在下列文件的物理结构中，（ ）最容易造成文件内容丢失

A、哈希分配　B、连续分配

C、链接分配　D、索引分配

2、不需要额外存储空间的磁盘空间的分配方式是（ ）

A、连续分配　B、隐式链接分配

C、索引分配　D、显式链接分配

3、（ ）分配方式无法快速读取文件的中间一块

A、连续　B、显式链接　C、隐式链接　D、索引

4、设某文件为显式链接文件，由5个逻辑记录组成，每个逻辑记录的大小与磁盘块大小相等，均为1KB字节，并依次存放在50、121、75、80、63号磁盘块上。若要存取文件的逻辑地址为6000处的信息，要访问的磁盘块分别是（ ）

A、其它　B、5　C、地址越界　D、63

试一试

- 5、若是一个磁盘容量是64MB，磁盘盘块大小为1KB，若是采用显式链接的方式，需要多大的FAT表；若是用索引结构，需要用几级索引，为什么？

- 6、设文件索引节点中有7个地址项，其中4个地址项是直接地址索引，2个地址项是一级间接地址索引，1个地址项是二级间接地址索引，每个地址项大小为4字节。若磁盘索引块和磁盘数据块大小均为256字节，则可表示的单个文件最大长度为多少？

- 7、为支持CD-ROM中视频文件的快速随机播放，播放性能最好的文件数据块组织方式是哪种？

8.2 文件存储空间的管理

8.2.1 空闲表法

- 基本思想

系统为外存所有空闲区建立一张空闲表，每个空闲区对应一个空闲表项。

序号	第一空闲盘块号	空闲盘块数
1	2	4
2	9	3
3	15	5
4	—	—

空闲盘块表

8.2 文件存储空间的管理

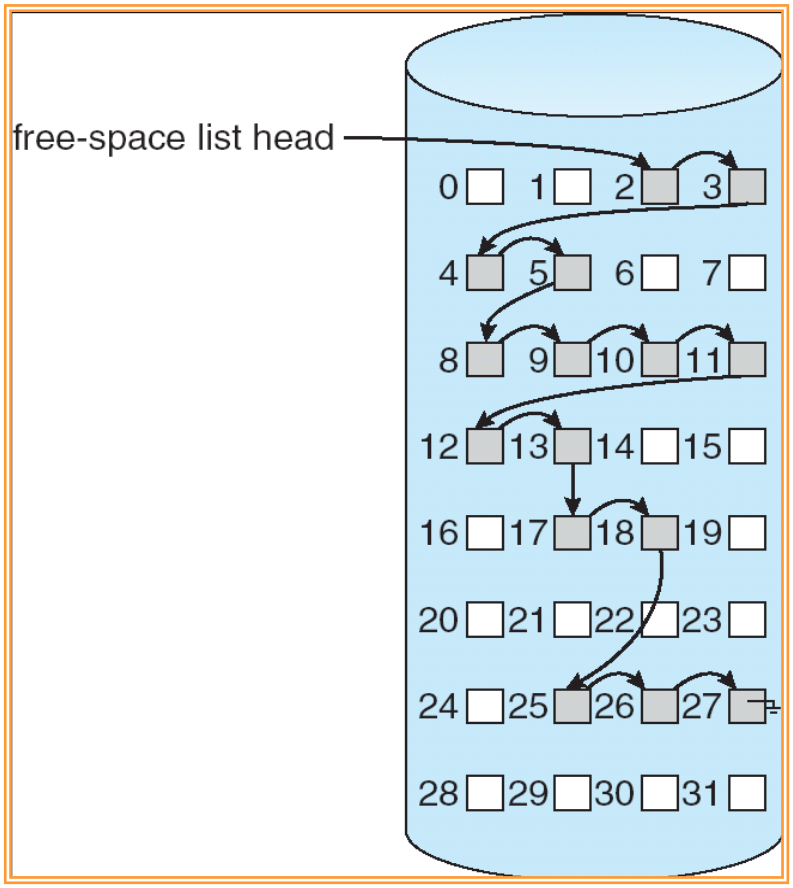
8.2.2 空闲链表法

- 基本思想

系统为外存所有空闲区拉成一张空闲盘区（块）表。

空闲盘块链：结点代表一个空闲盘块；

空闲盘区链：结点代表一个空闲盘区；



8.2 文件存储空间的管理

8.2.3 位示图法

- 基本思想
- 位示图利用二进制的一位来表示磁盘中一个盘块的使用情况。
  - ✓当其值为“0”时，表示对应的盘块空闲；
  - ✓为“1”时，表示已分配；

	0	1	2	3	4		…								15
0	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
1	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
2	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1
3	1	0	0	1	1	0	1	0	1	0	1	1	0	0	0	0
4	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
5	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
6	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1
7	1	0	0	1	1	0	1	0	1	0	1	1	0	0	0	0
	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1

Var bitmap: array[1…m] of integer_16 ;

8.2 文件存储空间的管理

8.2.3 位示图法

- 磁盘块号与位示图行号、列号计算公式
  - 磁盘块号b → 位示图行号i、列号j（b, i, j均从0开始计数）:

    i = b/n ;　　j = b % n;　　// n为位示图宽度，常为字长
  - 位示图行号i、列号j → 磁盘块号b:

    b = i*n + j ;

j

i	0	1	2	3	4	…									15
0	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
1	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
2	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1
3	1	0	0	1	1	0	1	0	1	0	1	1	0	0	0	0
4	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
5	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1
6	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1
7	1	0	0	1	1	0	1	0	1	0	1	1	0	0	0	0
	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1

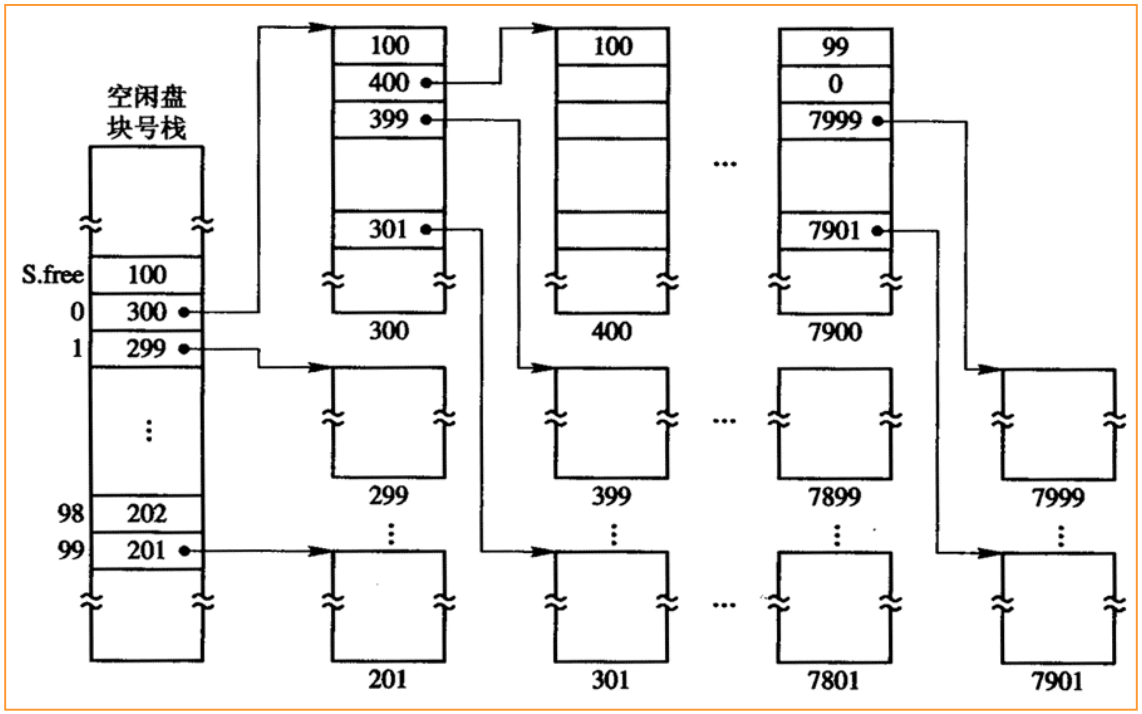
8.2 文件存储空间的管理

8.2.4 成组链接法

- 基本思想（UNIX）

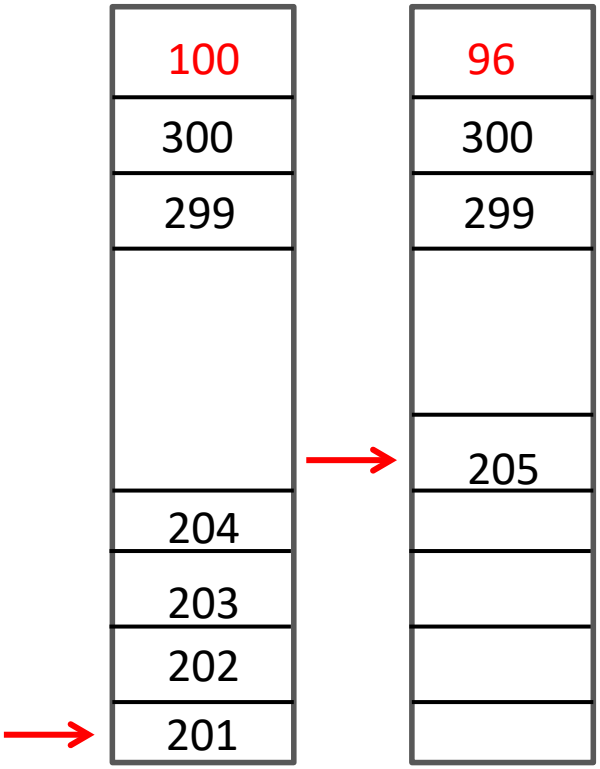
  利用链栈存储所有的空闲磁盘块号。

  链栈的每个结点存储一组（例如100个）磁盘块号，其中最后一个磁盘块号指向下一组所在的磁盘块。

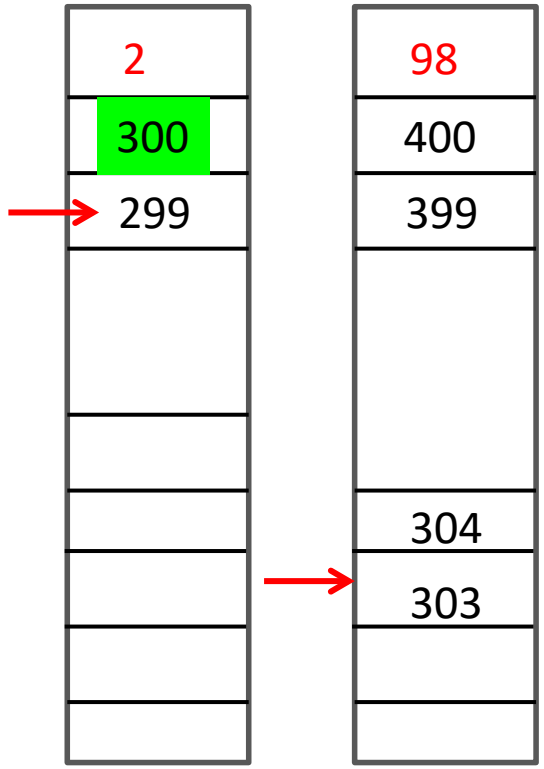


8.2 文件存储空间的管理

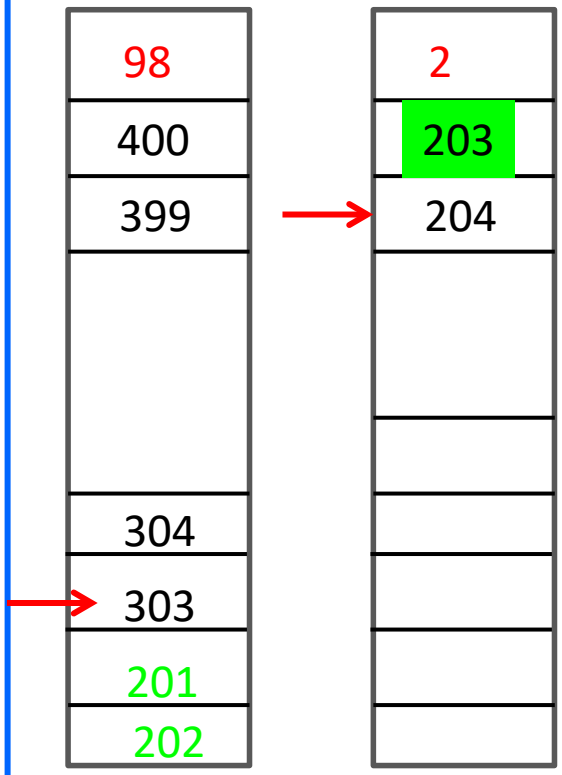
8.2.4 成组链接法



新建A文件
分配201-204四块



新建B文件
分配299-300两块，
复制#300块的内容，
再分配301，302



删除A文件
回收201-202两块，复制当前内容到#203块
再回收203，204

内存管理与磁盘管理比较

动态磁盘管理 ← 文件 → 静态物理结构

内存管理		
空闲表	分区表（连续）	顺序
空闲链表		链式（隐式）
FAT表		链式（显式）
位示图	位示图（离散）	顺序、链式、索引
成组链接		链式、索引

试一试

1、在位示图中的100号字节中的第1位（字节位）对应的物理块块号是（　）

A、801　B、806　C、101　D、798

2、负责管理磁盘空闲空间的模块在文件系统中是（　）

A、逻辑文件系统 B、磁盘系统

C、文件组织模块 D、基本文件系统

3、各种存储空间管理技术中，请对外存额外空间需求从大到小排列，对内存额外空间需求从大到小排列　最大的是（　）　最小的是（　）

A、空闲表　B、空闲链表　C、位示图　D、成组链接

试一试

- 4、一个空闲块位图开始时和磁盘分区首次初始化类似，比如：1000 0000 0000 0000（首块被根目录使用），系统总是从最小编号的盘块开始寻找空闲块，所以在有6块的文件A写入之后，该位图为1111 1110 0000 0000。请说明在离散分配方式中在完成如下附加动作之后位图的状态：

  a）写人有5块的文件B。

  b）删除文件A。

  c）写入有8块的文件C。

  d）删除文件B。

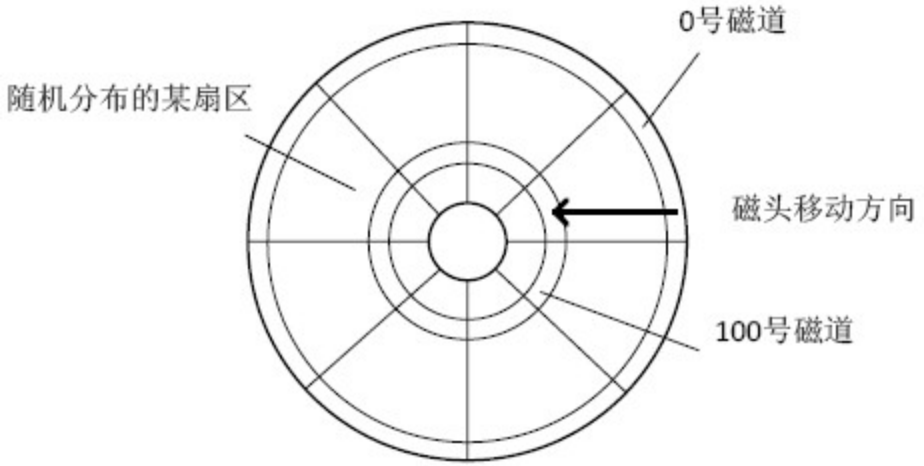
例题

假设计算机系统采用CSCAN（循环扫描）磁盘调度策略，使用2KB的内存空间记录16384个磁盘块的空闲状态。

（1）请说明在上述条件下如何进行磁盘块空闲状态的管理

（2）设某单面磁盘旋转速度为每分钟6000转，每个磁道有100个扇区，相邻磁道间的平均移动时间为1ms。若在某时刻，磁头位于100号磁道处，并沿着磁道号增大的方向移动，磁道号请求队列为50，90，30，120，对请求队列中的每个磁道需读取1个随机分布的扇区，则读完这4个扇区点共需要多少时间？要求给出计算过程。

（3）如果将磁盘替换为随机访问的FLASH半导体存储器（如U盘、SSD等），是否有比CSCAN更高效的磁盘调度策略？若有，给出磁盘调度策略的名称并说明理由；若无，说明理由。



(1) 用位图表示磁盘的空闲状态。每一位表示一个磁盘块的空闲状态，共需要 16 384 / 32=512 个字=512×4 个字节=2KB，正好可放在系统提供的内存中。

(2) 采用 CSCAN 调度算法，访问磁道的顺序和移动的磁道数如下表所示：

被访问的下一个磁道号	移动距离（磁道数）
120	20
30	90
50	20
90	40

移动的磁道数为 20+90+20+40=170，故总的移动磁道时间为 170ms。

由于转速为 6000r/m，则平均旋转延迟为 5ms，总的旋转延迟时间=20ms。

由于转速为 6000r/m，则读取一个磁道上一个扇区的平均读取时间为 0.1ms，总的读取扇区的时间平均读取时间为 0.1ms，总的读取扇区的时间为 0.4ms。

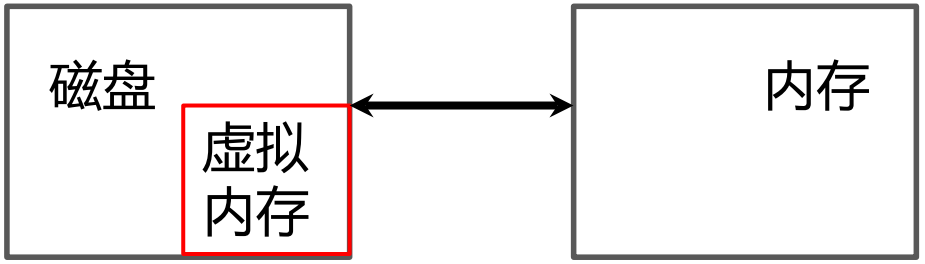
综上，读取上述磁道上所有扇区所花的总时间为 190.4ms。

(3) 采用 FCFS（先来先服务）调度策略更高效。因为 Flash 半导体存储器的物理结构不需要考虑寻道时间和旋转延迟，可直接按 I/O 请求的先后顺序服务。

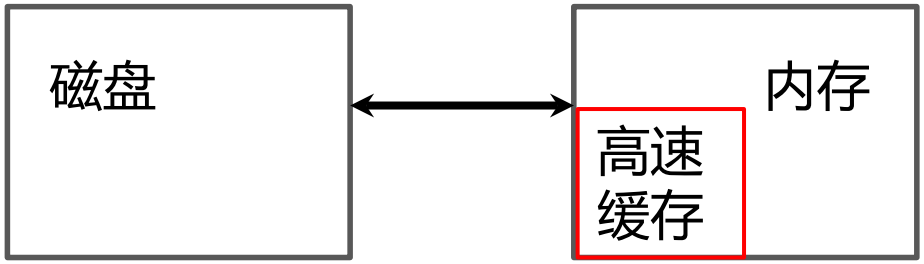
8.3 提高磁盘I/O速度的途径

8.3.1 磁盘高速缓存

- 虚拟存储器
  - ✓在磁盘中设置内存对换区
  - ✓时间换容量

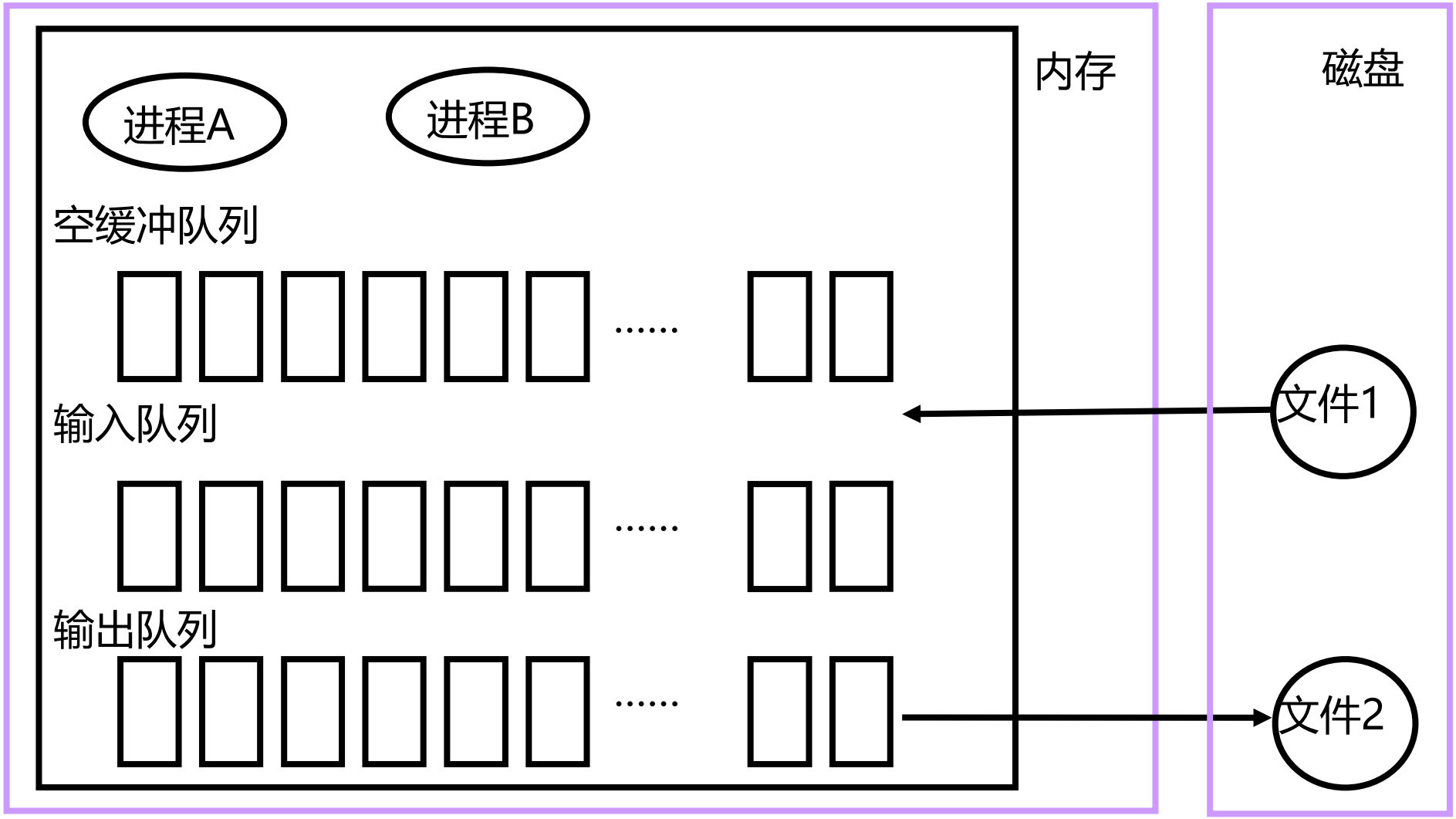


- 磁盘高速缓存
  - ✓在内存中设置磁盘缓冲区
  - ✓容量换时间



8.3 提高磁盘I/O速度的途径

8.3.1 磁盘高速缓存



8.3 提高磁盘I/O速度的途径

8.3.1 磁盘高速缓存

- 数据交付
  - ✓数据交付：直接传递数据
  - ✓指针交付
- 置换方法
  - ✓借鉴页面置换算法（LRU、NRU、LFU）
  - ✓访问频率低、可预见性、数据一致性
- 周期性写回
  - ✓文档保护方案：参考word等软件的临时保存方式

8.3 提高磁盘I/O速度的途径

8.3.2 其它一些方式

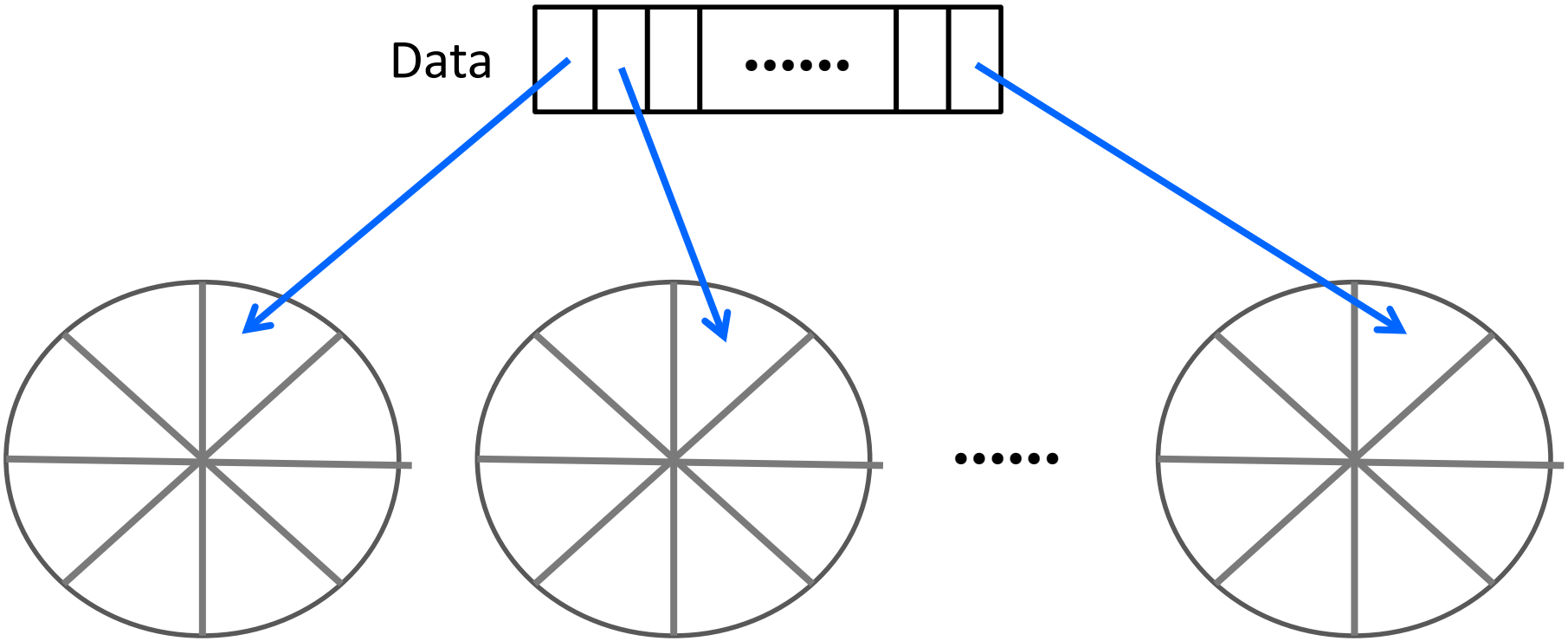
- 提前读
  - ✓减少启动I/O的次数
- 延迟写
  - ✓减少启动I/O的次数——与周期性写回矛盾?
- 优化物理块分布
  - ✓物理块分配尽量同磁道
  - ✓磁盘整理工具
- 虚拟盘
  - ✓RAM盘，如：虚拟光驱
  - ✓磁盘高速缓存——OS控制，RAM盘——用户控制

8.3 提高磁盘I/O速度的途径

8.3.3 廉价磁盘冗余阵列（RAID）

- 并行交叉存取

  将数据分成n份，分别存入不同磁盘的同一个扇区



RAID的主要技术之一，但并不是所有的RAID功能

8.3 提高磁盘I/O速度的途径

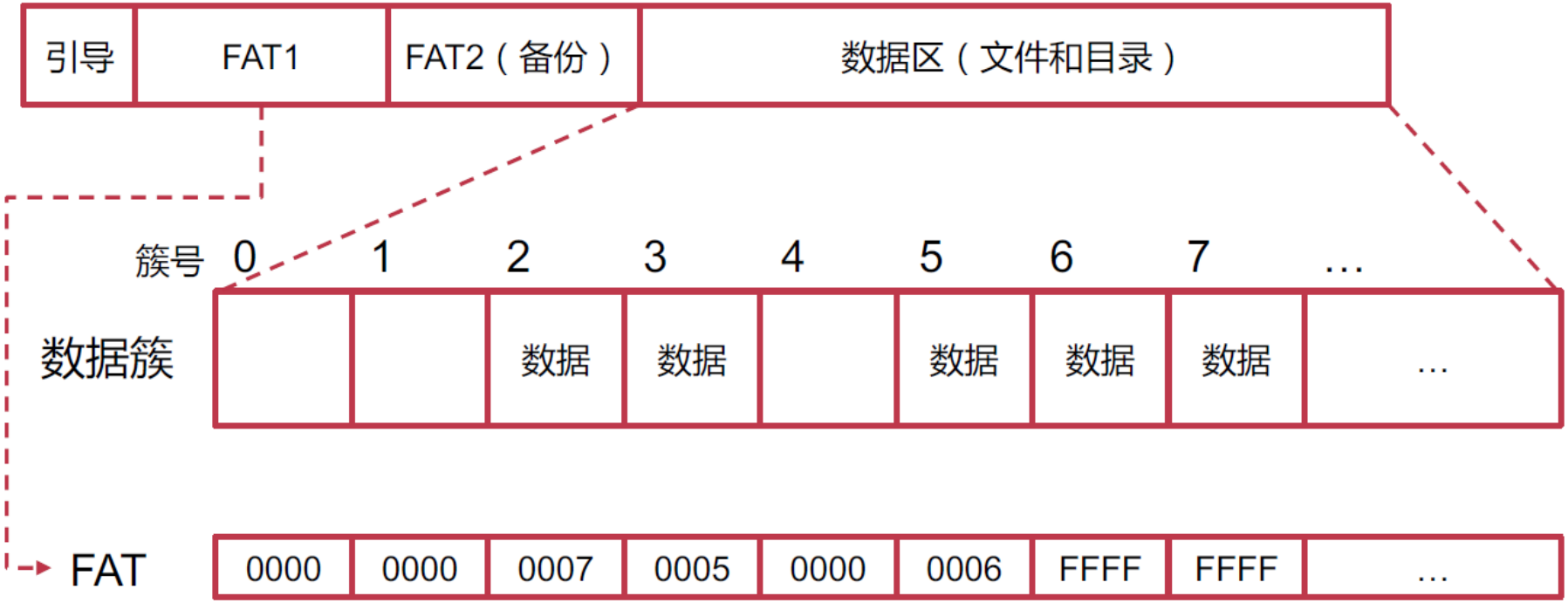
8.3.3 RAID分级

- RAID 0：仅提供并行交叉存取（只求速度）
- RAID 1：0+磁盘镜像（利用率50%）
- RAID 3：0+奇偶校验盘（利用率较高）
- RAID 5：0+独立数据通路
- RAID 6：3 和 5 的结合
- RAID 7：6 的改进

ORICO 睿阵提供 RAID 模式 0、1、3、5、10 和 Combine 等多种阵列存储模式。你可以根据工作需要组建不同目的性的存储方式，以利于你更快更好的完成所有任务。

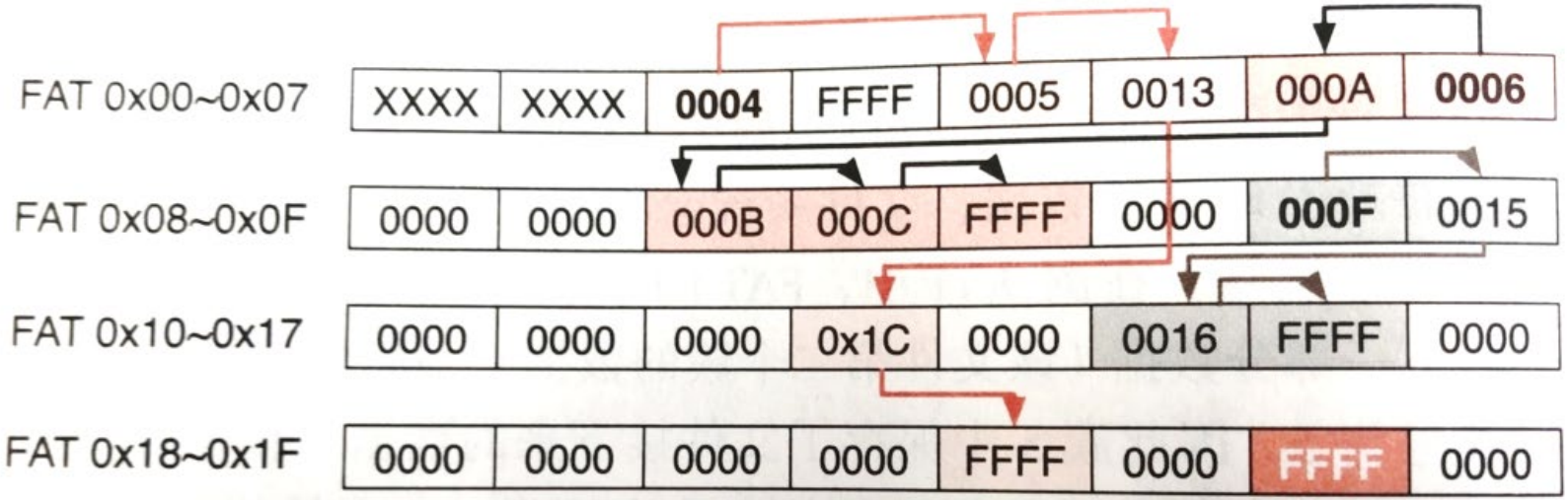
Raid模式	最少磁盘数	容量	安全性	速度
Raid 5(优选)	3			
Raid 0	2			
Raid 1	仅2			
Raid 3	3			
Raid 10	4			
Combine	2			
Clear	1			

补充：FAT32——磁盘布局



FAT为每个数据簇增加了一个**next**指针，让簇可以串联在一起

补充：FAT32——文件分配表



根文件夹的开始簇号为2，其内容保存在簇2、4、5、0x13、0x1C中。

文件DOCUM.DOC的开始簇号为7，其内容保存在簇7、6、0xA、0xB、0xC中。

文件夹MEDIA的开始簇号为0xE，其内容保存在簇0xE、0xF、0x15、0x16中。

文件SCORE.TXT的开始簇号为0x1E，其内容保存在簇0x1E中。

补充：FAT32——目录项

两个目录项组成的长文件名

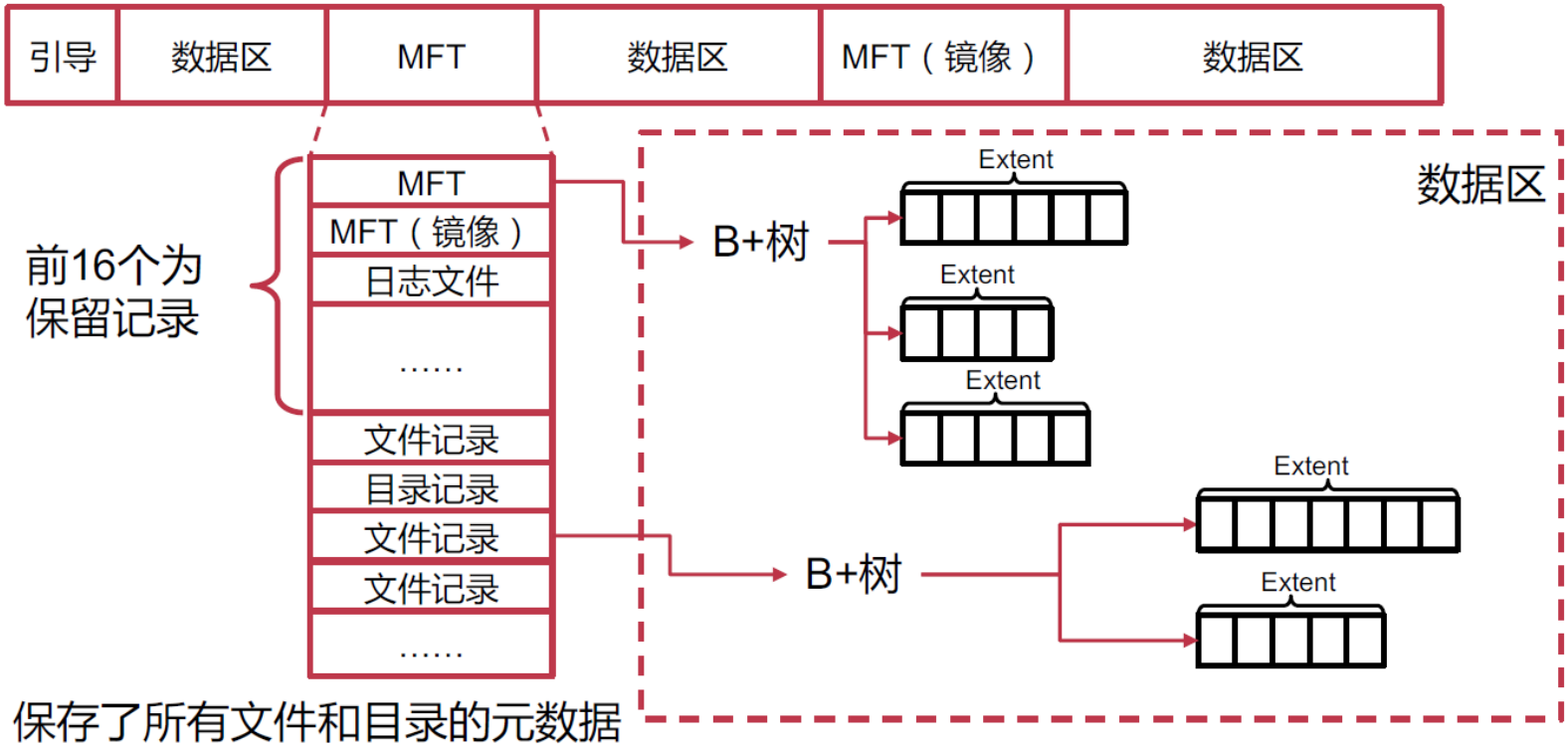
0x42	y	s	t	e	m	0x0F	0x00	校验	.
t	e	x	0x0000	0xFFFF	0x0000	0xFFFF			0xFFFF
0x01	O	S	B	o	o	0x0F	0x00	校验	k
	F	i	l	e	0x0000				S

短文件名

O	S	B	O	O	K	~	1	T	E	X	0x20	NT	创建时间
创建日期		访问日期		0x0000		修改时间		修改日期		起始簇		文件大小	

长文件名：在使用短文件名的基础上，用多个额外的目录项来保存长文件名。该类型用0x0F表示。
一个额外目录项可以保存13个Unicode字符（上图中为 OSBook File System.tex）
FAT32限制文件名的长度不能超过255（不超过20个目录项）

补充：NTFS——磁盘布局



MFT（Master File Table）：主文件表，微软称其为关系型数据库，每一行对应一个文件，每一列为这个文件的某个元数据（属性）。

补充：NTFS——MFT

序号	文件名	说明
0	$MFT	主文件表
1	$MFTMirr	主文件表镜像
2	$LogFile	日志文件
3	$Volume	卷文件
4	$AttrDef	属性定义列表
5	$Root	根目录
6	$Bitmap	位图文件
7	$Boot	引导文件
8	$BadClus	坏簇文件
9	$Secure	安全文件
10	$UpCase	大写表
11	$Extend	扩展元数据目录
12	$Extend\$Reparse	重解析点文件
13	$Extend\$UsnJrnl	变更日志文件
14	$Extend\$Quota	配额管理文件
15	$Extend\$ObjId	对象ID文件

$MFT：元数据文件保存了MFT的所有内容（即MFT记录），因此MFT能够“自己管理自己”。引导区只记录MFT的开始16条记录。通过到找$MFT文件所有数据保存位置，才能找到剩余所有记录。

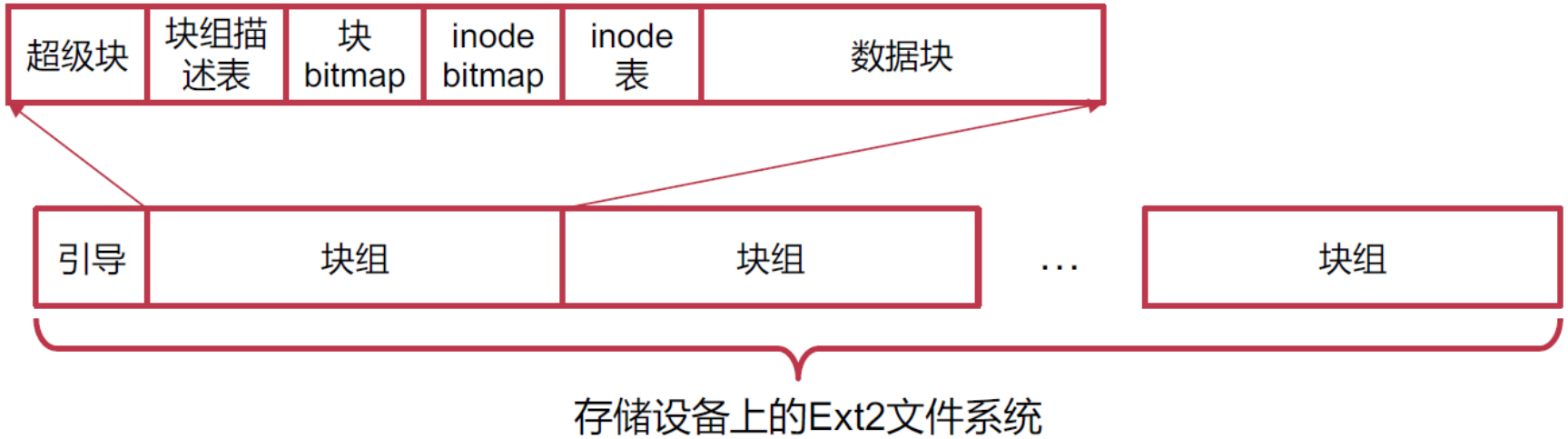
- 非常驻文件（大文件/目录）
  - 数据区的B+树和区段
- 常驻文件（小文件/目录）
  - 大小不超过MFT记录的最大值（1KB）
  - 内嵌在MFT中保存（在数据属性中）
- 目录项
  - 包含文件名、文件ID（在MFT中的序号）

补充：Ext2—存储布局

将磁盘分为多个块组，每个块组中都有超级块，互为备份

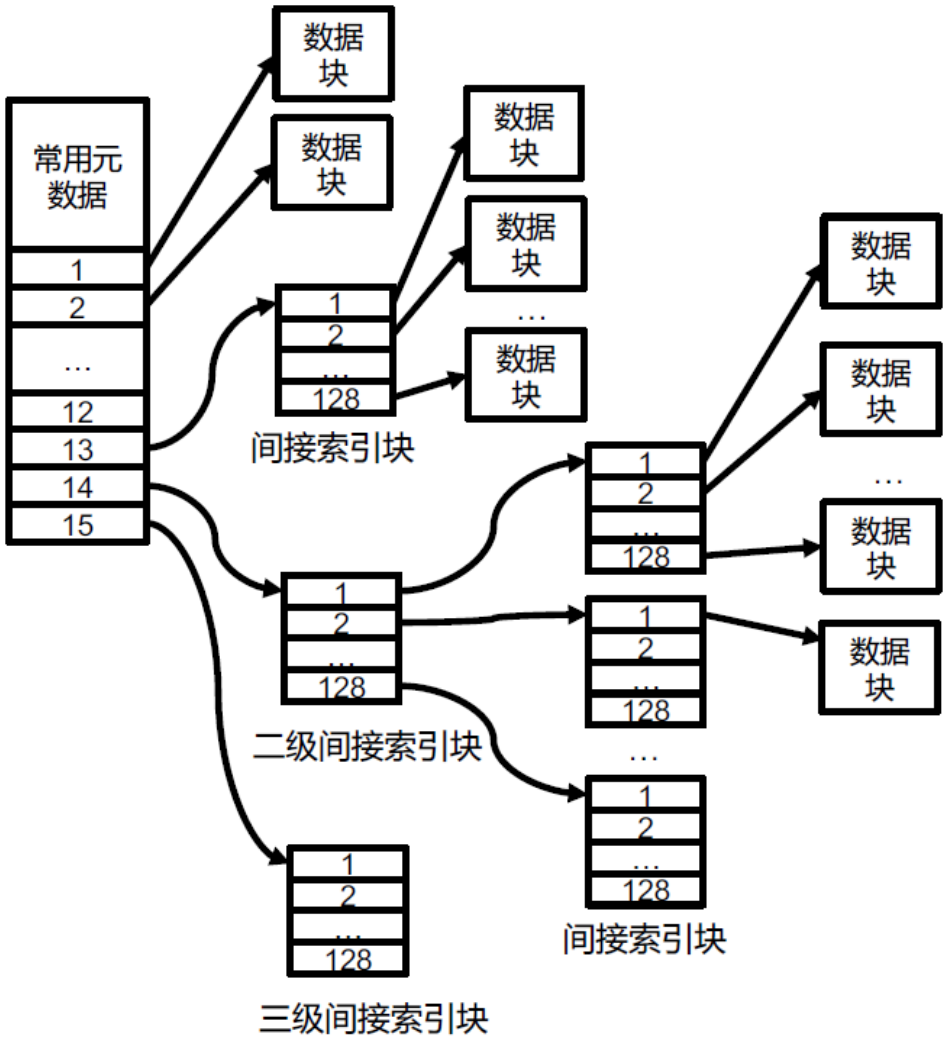
超级块（Super Block）记录了整个文件系统的元数据

块组描述表记录了快速中各个区域的位置和大小

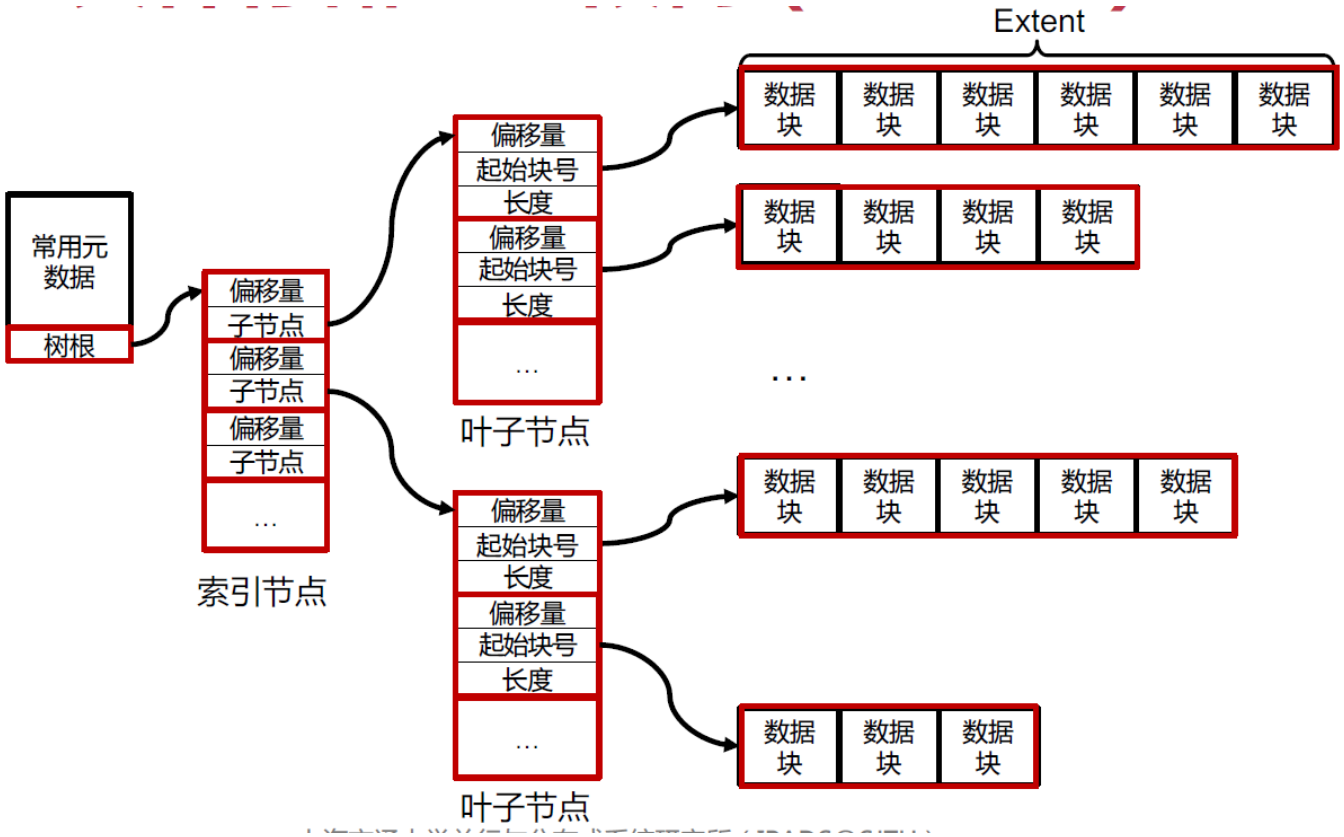


补充：Ext2—inode

- 常用的元数据
  - 文件类型
  - 文件大小
  - 链接数
  - 文件权限
  - 拥有用户/组
  - 时间（创建、修改、访问时间）



补充：Ext4—区段树（Extent）

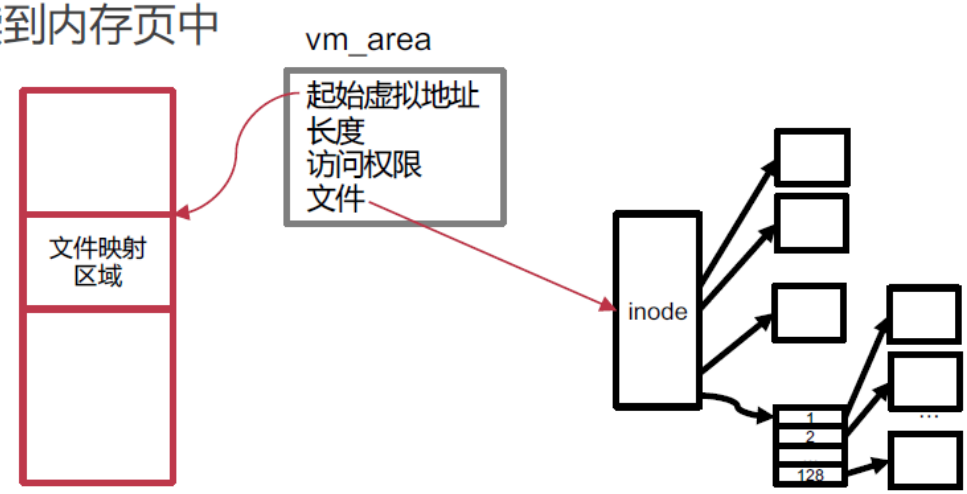


区段： 是由物理上连续的多个数据块组成
一个区段内的数据可以连续访问，无需按4KB数据块访问

补充：文件内存映射

- mmap可将文件映射到虚拟内存空间中
  1. mmap时分配虚拟地址，并标记此段虚拟地址与该文件的inode绑定
  2. 访问mmap返回的虚拟地址时，触发缺页中断（page fault）
  3. 缺页中断处理函数，通过虚拟地址，找到该文件的inode
  4. 从磁盘中将inode中对应的数据读到内存页中
  5. 将内存页映射添加到页表中

```
fd = open("/OS/考试", O_RDWR);
addr = mmap(NULL, length, PROT_WRITE,
            MAP_SHARED, fd, 0);
memset(addr, 0, length);
```



课后作业

1、一个文件有20个磁盘块（块号：0-19），假设文件控制块在内存（如果文件采用索引分配，索引表不在内存）。在下列情况下，请计算在连续分配，链接分配，单级索引分配三种分配方式下，分别需要多少次磁盘I/O操作？（每读入或写出一个磁盘块需要一次磁盘I/O操作，另外，假设在连续分配方式下，文件头部无空闲的磁盘块，但文件尾部有空闲的磁盘块。

1）在文件开始处删除一个磁盘块；

2）在文件第15块前添加一个磁盘块并写入内容；

3）在文件结尾处删除一个磁盘块；

4）在文件结尾处增加一个磁盘块并写入内容。

2、某文件系统空间的最大容量为4TB(1T=240),以磁盘块为基本分配单位,磁盘块大小为1KB。文件控制块(FCB)包含一个512B的索引表区。请回答下列问题。

(1)假设索引表区仅采用直接索引结构,索引表区存放文件占用的磁盘块号。索引表项中块号最少占多少字节?可支持的单个文件最大长度是多少字节?

(2)假设索引表区采用如下结构:第0~7字节采用(起始块号,块数)格式表示文件创建时预分配的连续存储空间,其中起始块号占6B,块数占2B;剩余504字节采用直接索引结构,一个索引项占6B,则可支持的单个文件最大长度是多少字节?为了使单个文件的长度达到最大,请指出起始块号和块数分别所占字节数的合理值并说明理由。

课后作业

3、某文件系统为一级目录结构，文件的数据一次性写入磁盘，已写入的文件不可修改，但可多次创建新文件。请回答如下问题。

（1）在连续、链式、索引三种文件的数据块组织方式中，哪种更合适？说明理由。为定位文件数据块，需要FCB中设计哪些相关描述字段？

（2）为快速找到文件，对于FCB，是集中存储好，还是与对应的文件数据块连续存储好？说明理由。