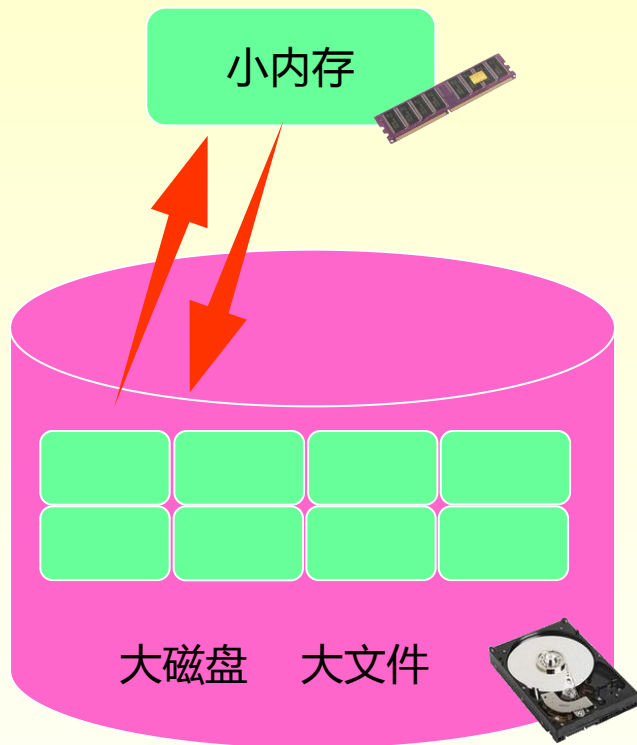




本节主题：

磁盘排序

# 磁盘排序面临的挑战



任务管理器

文件(F) 选项(O) 查看(V)

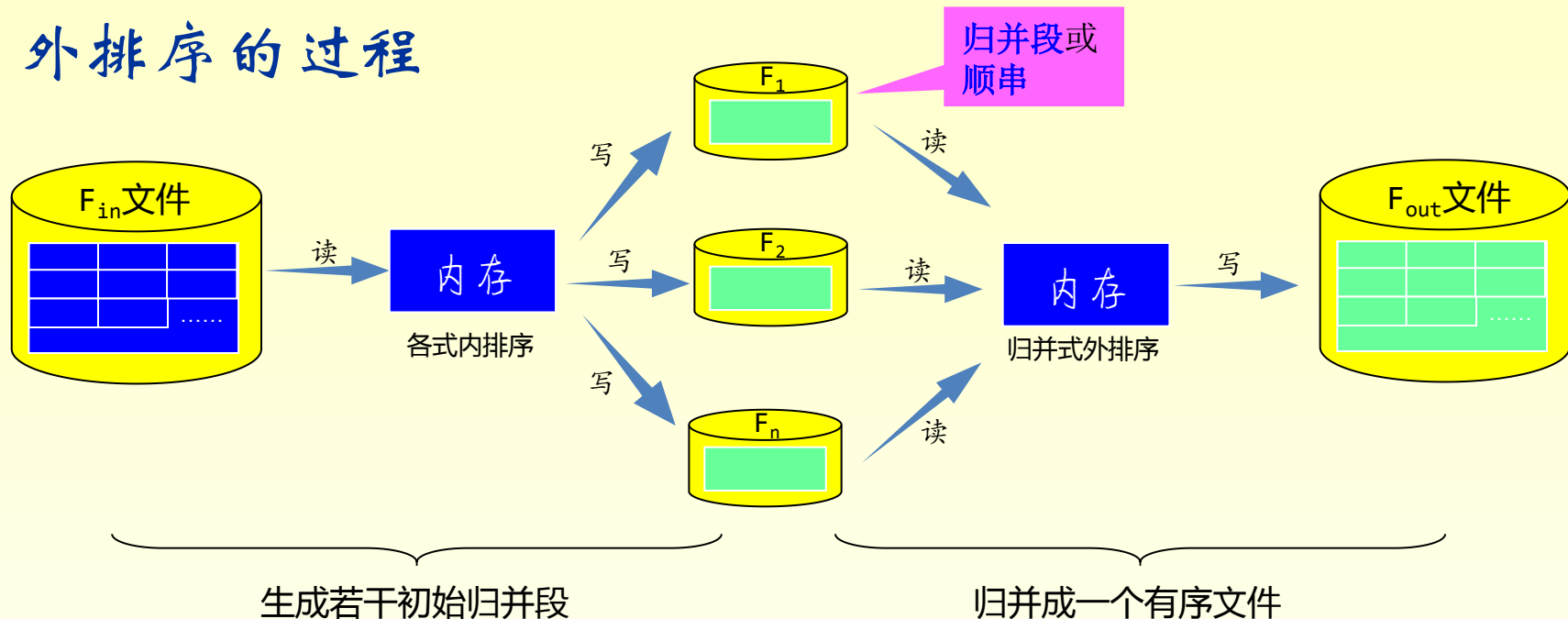
进程 性能 应用历史记录 启动 用户 详细信息 服务

名称	状态	6% CPU	56% 内存	21% 磁盘	0% 网络
应用 (5)					
Windows 资源管理器		0.1%	21.2 MB	0.1 MB/秒	0 Mbps
WPS Presentation (32 位)		0%	63.6 MB	0 MB/秒	0 Mbps
截图工具		0.4%	4.7 MB	0.1 MB/秒	0 Mbps
任务管理器		0.1%	12.8 MB	0.1 MB/秒	0 Mbps
搜狗高速浏览器 (32 位)		0%	32.7 MB	0 MB/秒	0 Mbps
后台进程 (62)					
ABC SafePop (32 位)		0%	0.8 MB	0 MB/秒	0 Mbps
Alipay security business...		0%	2.7 MB	0 MB/秒	0 Mbps
Alipay Security Server (3...		0%	1.0 MB	0 MB/秒	0 Mbps

简略信息(D) 结束任务(E)

在大文件、小内存的限制下，  
尽可能少地读写慢速的磁盘！

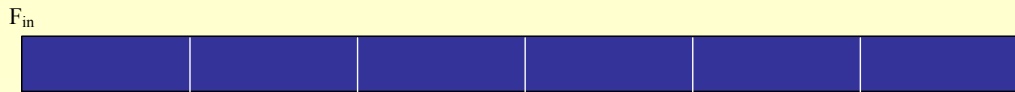
# 外排序的过程



- 按内存(缓冲区)的大小, 将 $n$ 个记录的数据文件分成若干个子段或子文件, 依次读入内存并选择内部排序方法进行排序;
- 将排好序的有序子文件重新写入到外存。

- 采用归并的办法对归并段进行逐趟归并, 使归并段的长度逐渐增大, 直到最后合并成只有一个归并段的文件, 得到排好序的文件。

# 初始归并段示例



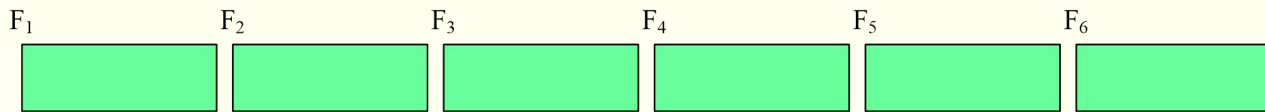
文件  $F_{in}$

4500个记录： $A_1, A_2, \dots, A_{4500}$

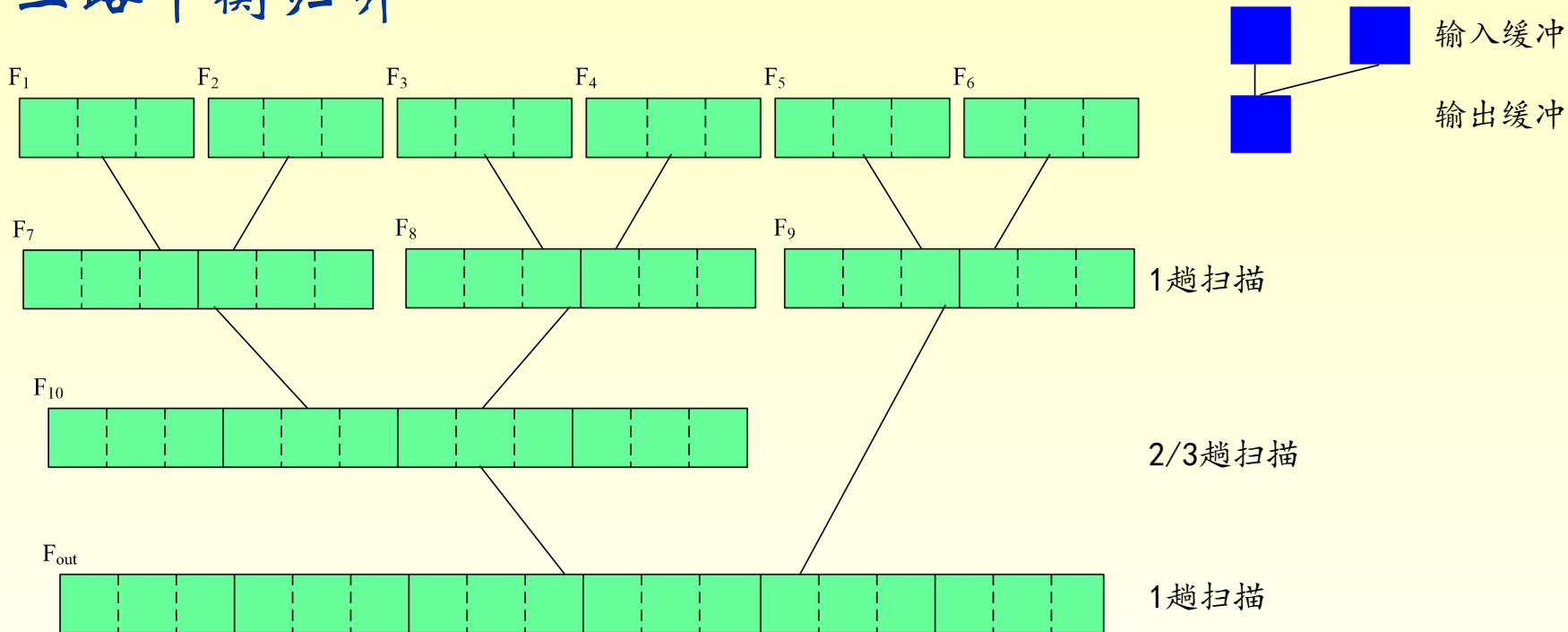


$M_1$   $M_2$   $M_3$

- 可使用的内存空间
  - 仅够750个记录进行排序
- 页块大小
  - 相当于250个记录大小
  - 磁盘每次读/写的单位
  - 一个物理块对应250个逻辑记录



# 二路平衡归并



小结：

- 等距初始归并段
- 二路归并

评价

- 成本高

改进方向

- K路(平衡)归并
- 初始时，产生尽可能长的初始归并段