

# 第11章 外排序

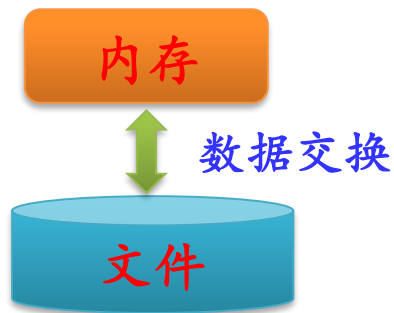
## 11.1 外排序概述

## 11.2 磁盘排序

# 11.1 外排序概述

## 1、什么是外排序

**外排序**是指数据存放在外存中，数据排序时涉及内、外存数据交换的排序方法。



存储在外存上的数据以文件为基本单位。

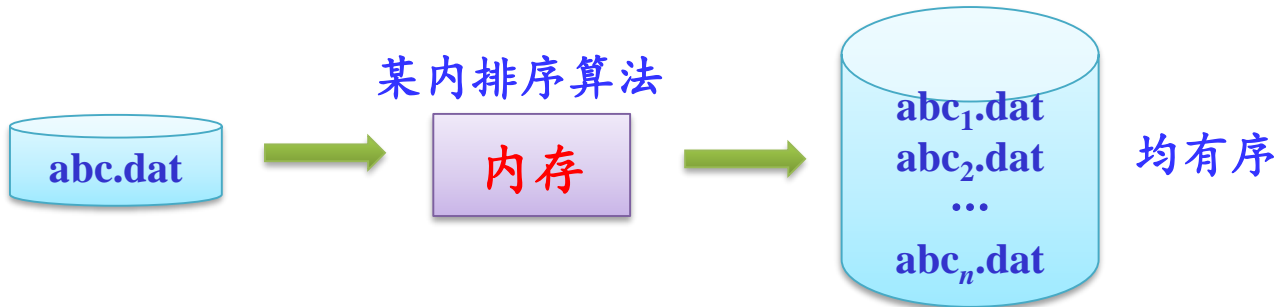
## 2、外排序的基本方法

外排序的基本方法是归并排序法。它分为以下两个步骤：

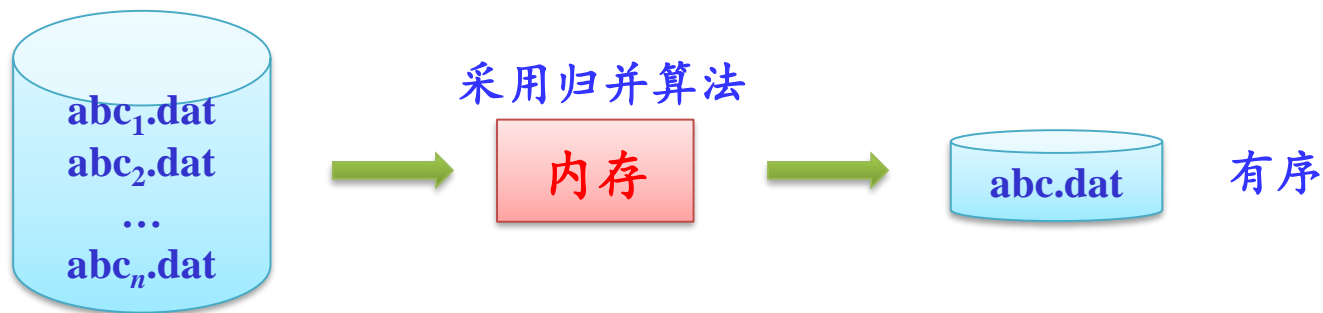
**(1) 生成若干初始归并段（顺串）：**这一过程也称为文件预处理。  
一种常规的方法如下：

① 把含有 $n$ 个记录的文件，按内存大小 $w$ 分成若干长度为 $w$ 的子文件（归并段）；

② 分别将各子文件（归并段）调入内存，采用有效的内排序方法排序后送回外存。产生 $\lceil n/w \rceil$ 个初始归并段。



(2) **多路归并**：对这些初始归并段进行多遍归并，使得有序的归并段逐渐扩大，最后在外存上形成整个文件的单一归并段，也就完成了这个文件的外排序。



## 示例

文件abc.dat:

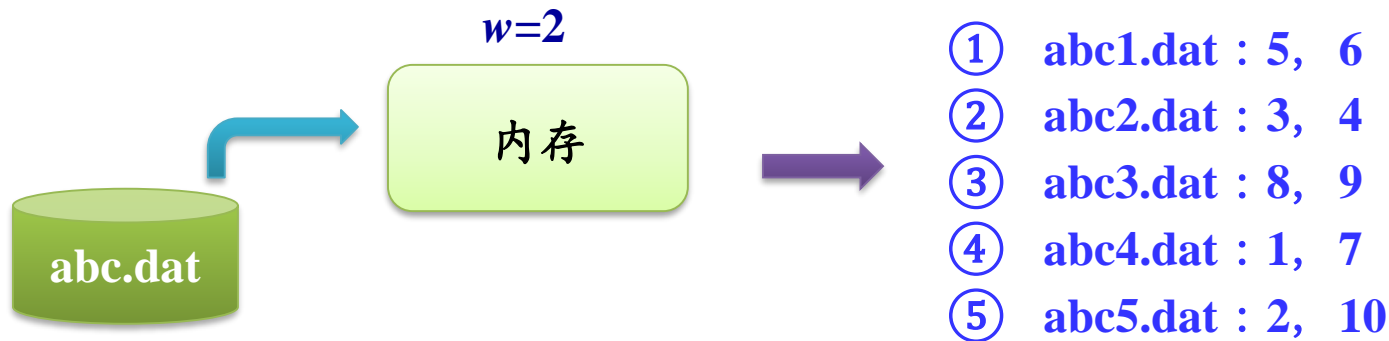
5, 6, 3, 4, 9, 8, 1, 7, 10, 2

← 递增排序

应用程序可用的内存空间大小 $w=2$ 。

## 外排序过程:

(1) 生成5个初始归并段



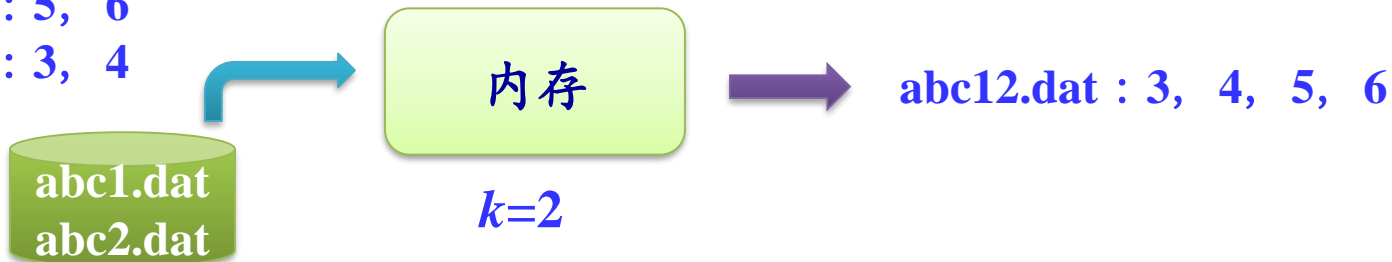
5, 6, 3, 4, 9, 8, 1, 7, 10, 2

## 外排序过程:

### (2) 多路归并: $w=2 \Rightarrow$ 2路归并 ( $k=2$ )

• abc1.dat : 5, 6

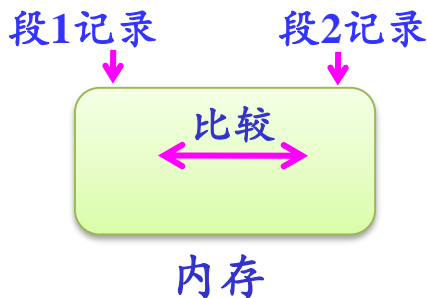
• abc2.dat : 3, 4



• abc1.dat : 5    6     $\infty$

• abc2.dat : 3    4     $\infty$

↓ 大致分析



结果段  
abc12.dat :

$u$  ( $u=4$ ) 个记录需要进行  $u-1$  次操作 (不考虑  $\infty$ )

$k$  ( $k=2$ ) 路归并每次需要  $k-1$  次关键字比较

} 总的关键字比较次数:  $(u-1)(k-1)$

- abc3.dat : 8, 9

- abc4.dat : 1, 7



$w=2$

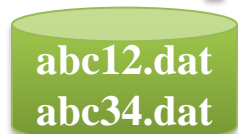


abc34.dat : 1, 7, 8, 9



● abc12.dat : 3, 4, 5, 6

● abc34.dat : 1, 7, 8, 9



$k=2$



abc1234.dat : 1, 3, 4,  
5, 6, 7, 8, 9

● abc1234.dat : 1, 3, 4,  
5, 6, 7, 8, 9

● abc5.dat : 2, 10

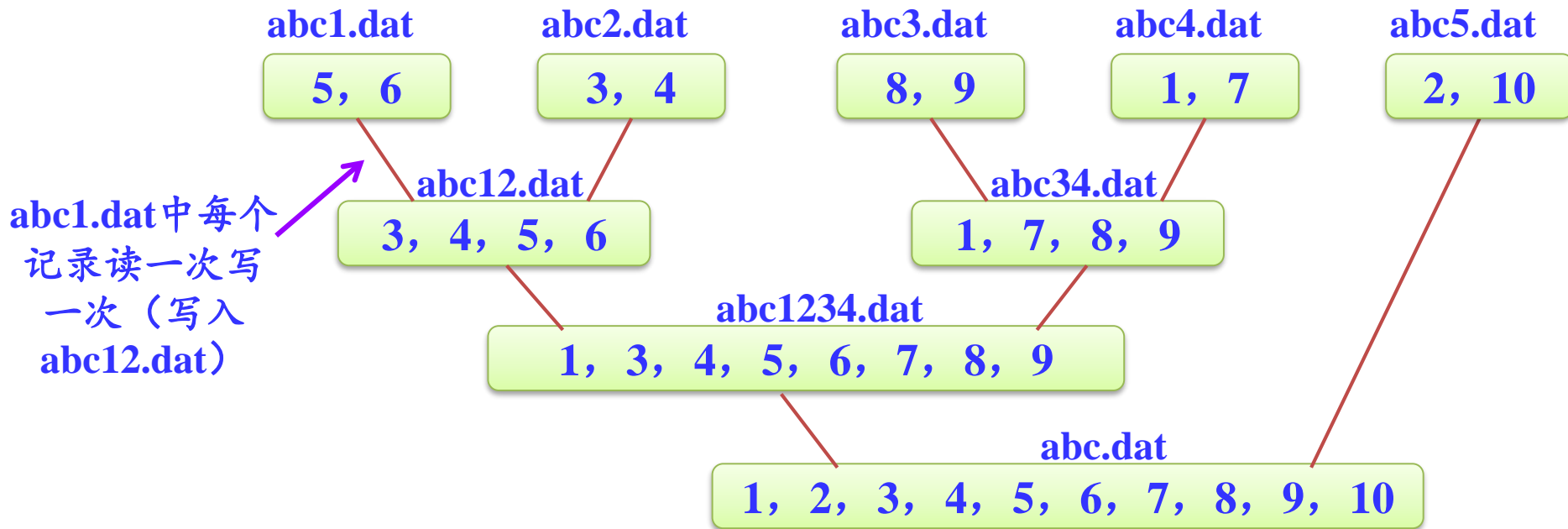


$k=2$



abc.dat : 1, 2, 3, 4,  
5, 6, 7, 8, 9, 10

## 归并过程对应的归并树



## 归并过程的性能:

- 记录读写次数。
- 内存中归并时需要关键字比较次数。

外排序方法与各种外存设备的特征有关。

外存设备大体上可分为两类：

- 顺序存取设备，例如磁带。
- 直接存取设备，例如磁盘。后面主要介绍磁盘排序。

## 思考题

外排序为什么不能直接采用前面介绍的基本内排序方法，如快速排序方法？

——本讲完——