

第4章 存储管理

- 1 概述
- 2 连续空间分配
- 3 非连续空间分配
- 4 虚存

研究三方面的问题：

- 取 (fetch) → 请调、预调
- 放 (placement) → 连续、非连续
- 替换 (replacement)

4.1 概述

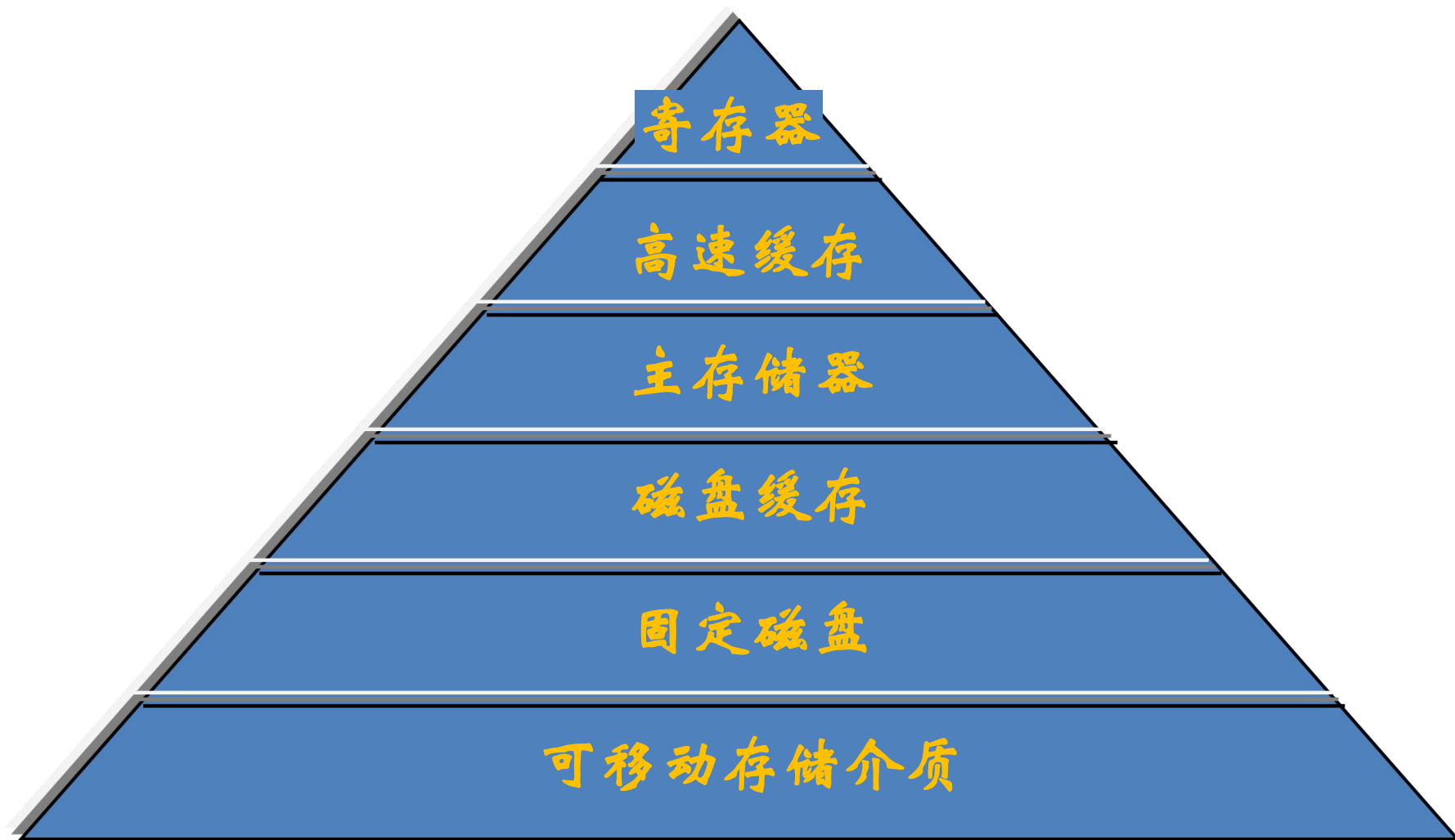
- 存储体系
 - “金字塔”问题
- 存储管理的目的及任务
- 内存的分配方式
 - 静态分配、动态分配
- 内存的管理方法
 - 分区管理、分页管理、分段管理等
- 相关概念及术语



4.1.1 存储管理的目的及任务

- 内存管理的相关内容
 - 内存管理方法
 - 内存分配和释放算法
 - 虚拟存储器的管理
 - 控制内存和外存之间数据流动方法
 - 地址变换技术
 - 内存数据保护与共享技术
- 存储管理的功能
 - 主存的分配和回收
 - 提高主存利用率：多道程序动态的共享主存
 - “扩充”主存容量：虚拟技术及自动覆盖技术
 - 存储保护
 - 内存共享

存储器的层次

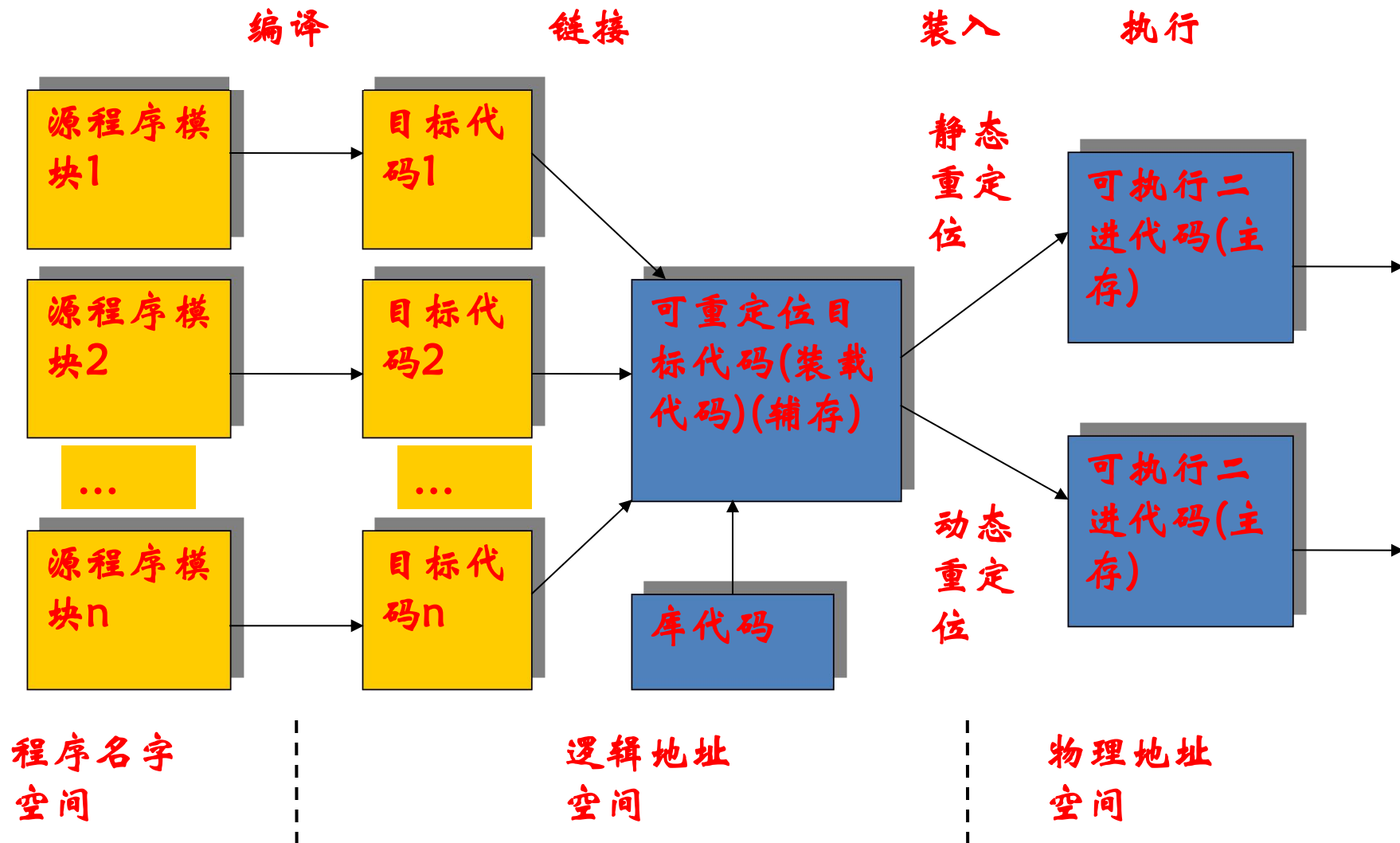


地址转换与存储保护

- 地址映射（重定位） **Relcation**
- 绝对地址与相对地址



程序的编译、链接、装入和执行



为保证cpu 执行指令时可正确访问存储单元，需将用户程序中的逻辑地址转换为机器直接寻址的物理地址的过程叫地址映射

表示：A：逻辑地址空间；M：内存空间

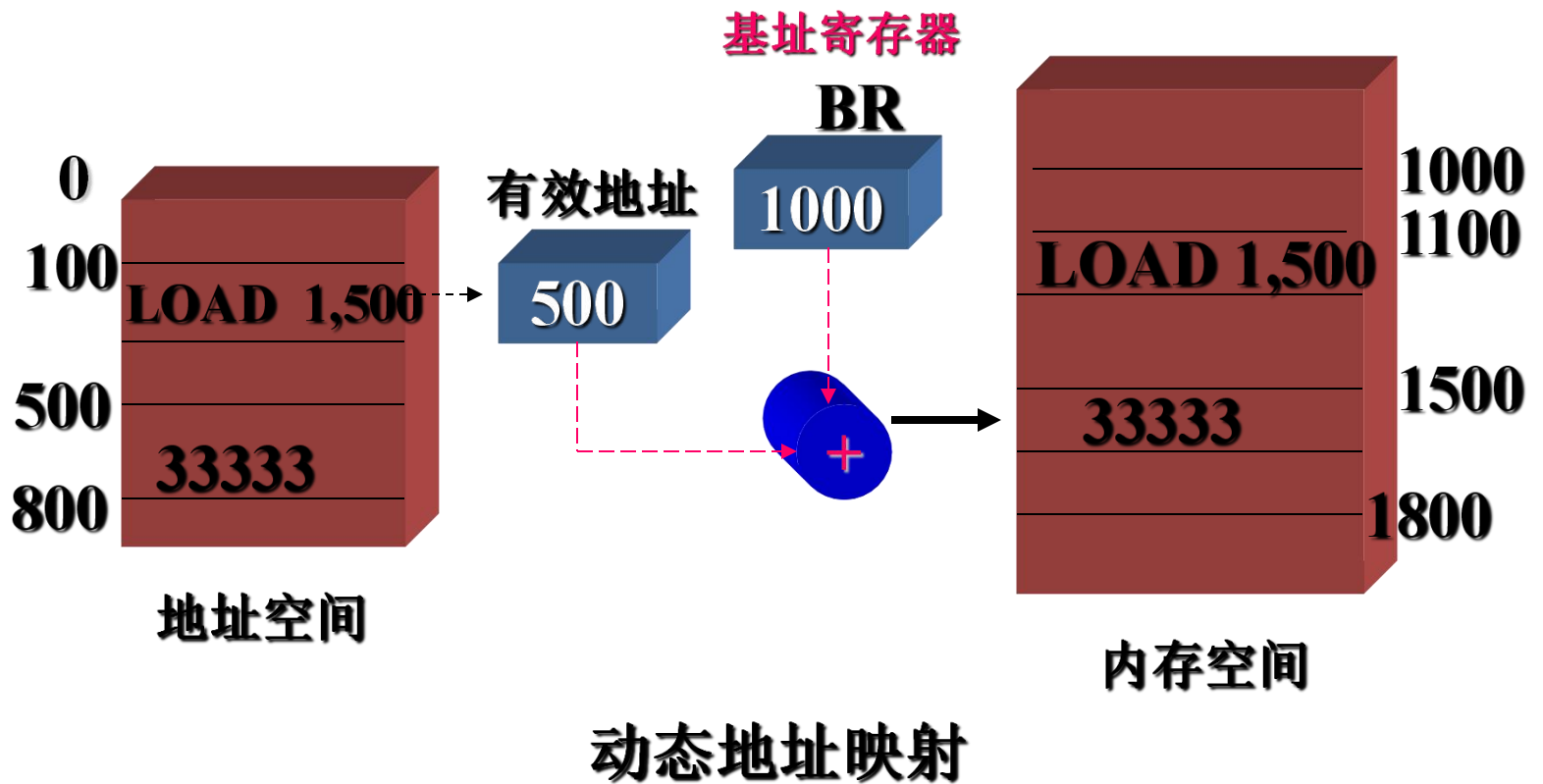
$$F = A \rightarrow M$$

静态重定位

- 一次性实现逻辑地址 \rightarrow 物理地址的转换，以后不再转换
- $f(a) = B + a$
 - $f(a)$: a 的物理地址
 - B : 起始地址
 - a 地址空间中的逻辑地址



动态重定位



4.2 连续空间分配

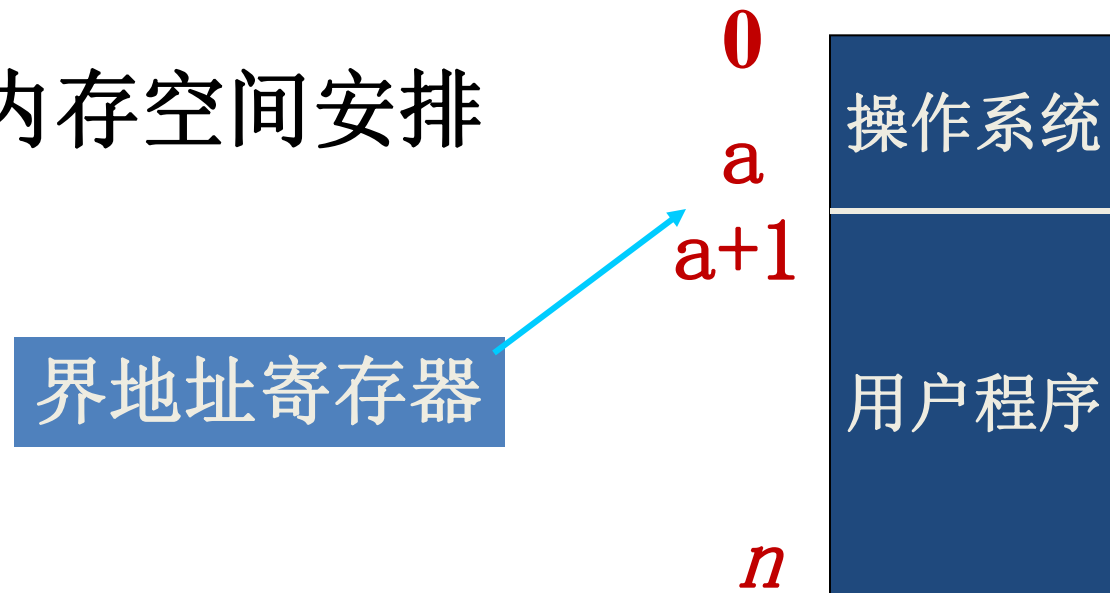
- 1 单道连续分配
- 2 多道固定分配
- 3 多道连续可变分配
- 4 伙伴系统

4.2.1 单道连续分配

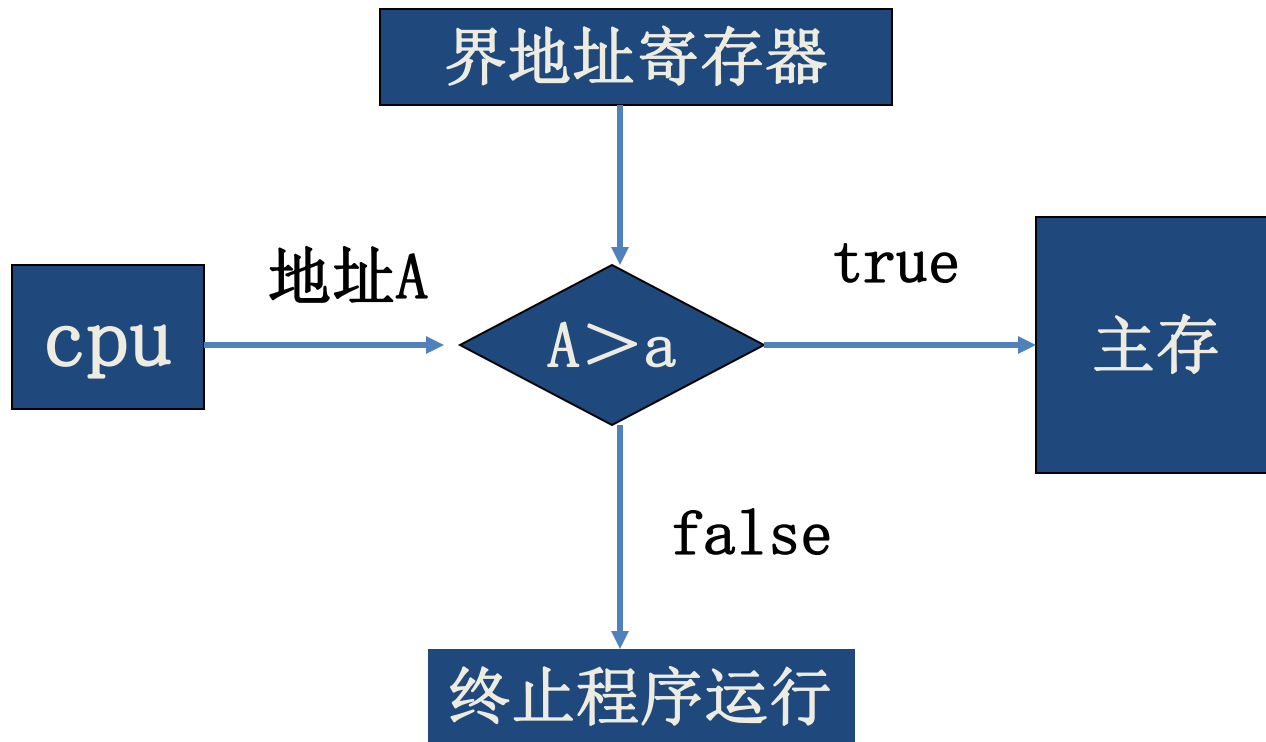
特点：任一时刻内存只有一道作业，该作业连续存放于内存中。

一、管理方法

内存空间安排

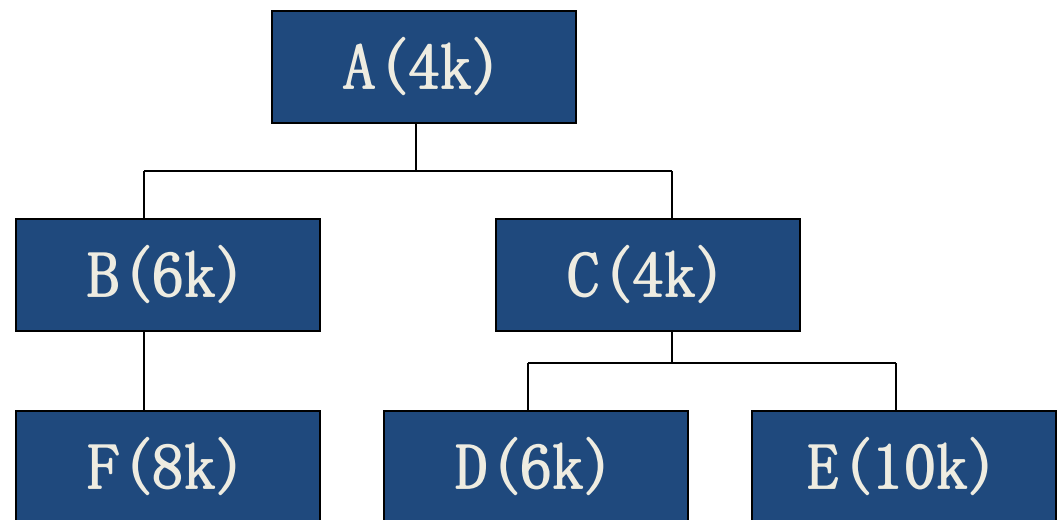
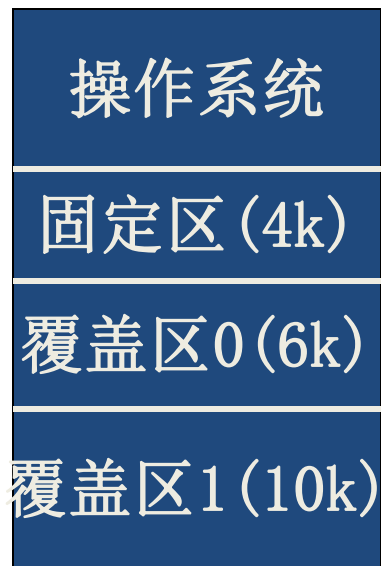


越界检查机构：用户程序每访问一次主存，越界检查机构将访问的地址与界地址寄存器中的值比较。若越界，则终止其执行。



二、覆盖（overlap）

因内存小于作业的程序空间而引入覆盖，将用户空间划分成一个固定区和多个覆盖区。主程序放固定区，依次调用的子程序则放在同一个覆盖区。操作系统提供覆盖系统调用函数，由用户编程序时考虑调用。



覆盖：组成覆盖段的过程

覆盖结构：

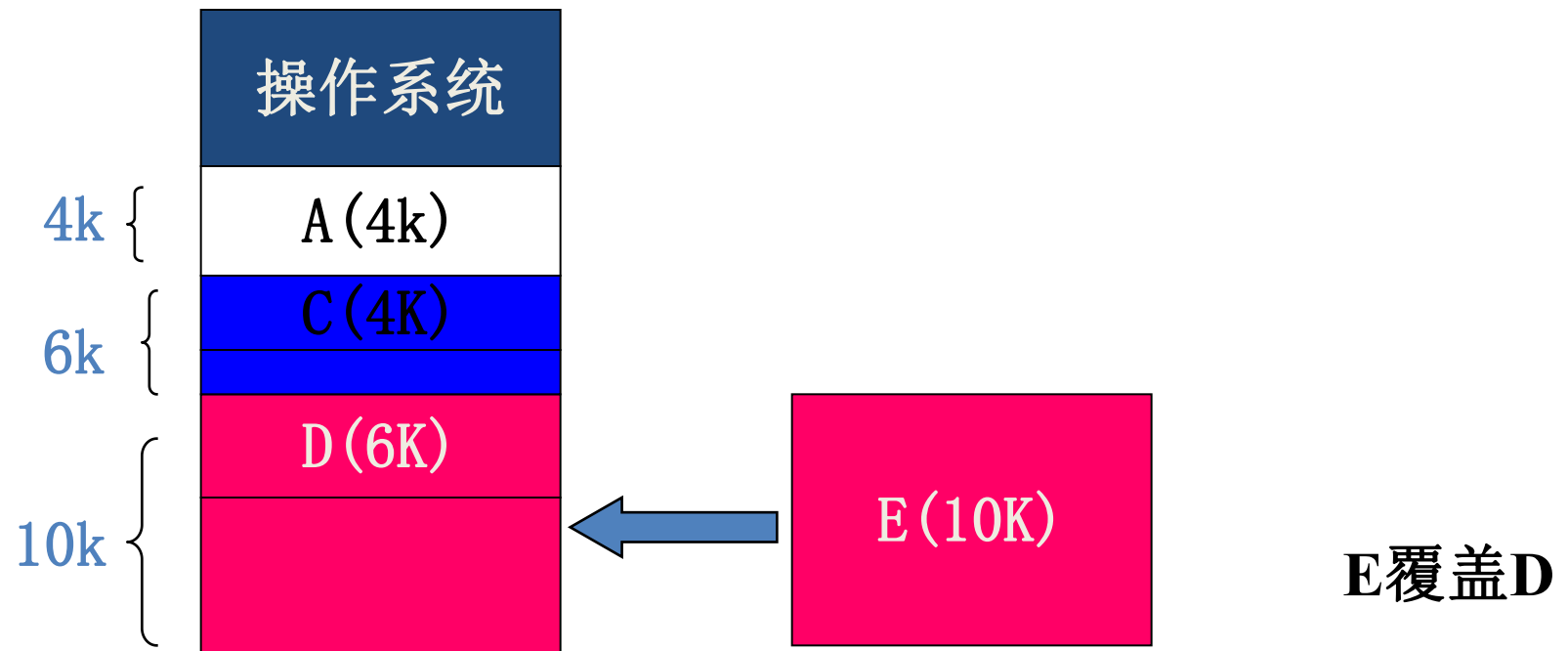
用 (i, j) 表示。 i 是覆盖段号； j 是覆盖段中覆盖号

覆盖段0： B、 C

B(0,0) C(0,1)

覆盖段1： F、 D、 E

F(1,0) D(1,1) E(1,2)



三、交换

基本思想：将处于等待状态(等I/O)或就绪状态的进程从主存换出到辅存，把将要执行的进程移入主存。

当一个作业正在进行I/O操作时，不能交换
引入系统区I/O缓冲区后，作业交换不受限制

优点：主存空间扩展，利用率提高
缺点：交换要花费较长的时间

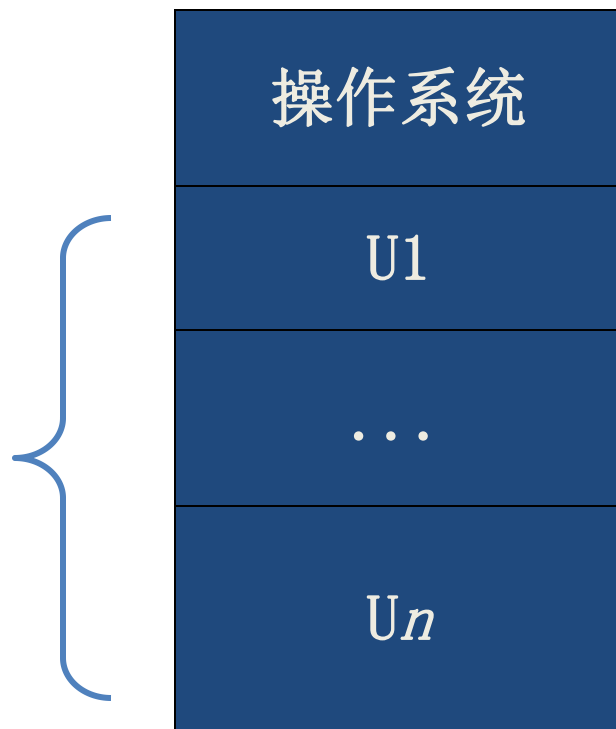
4.2.2 多道固定分区法

特点：任一时刻内存可有多道作业，每道作业连续存放于内存。

一、管理方法

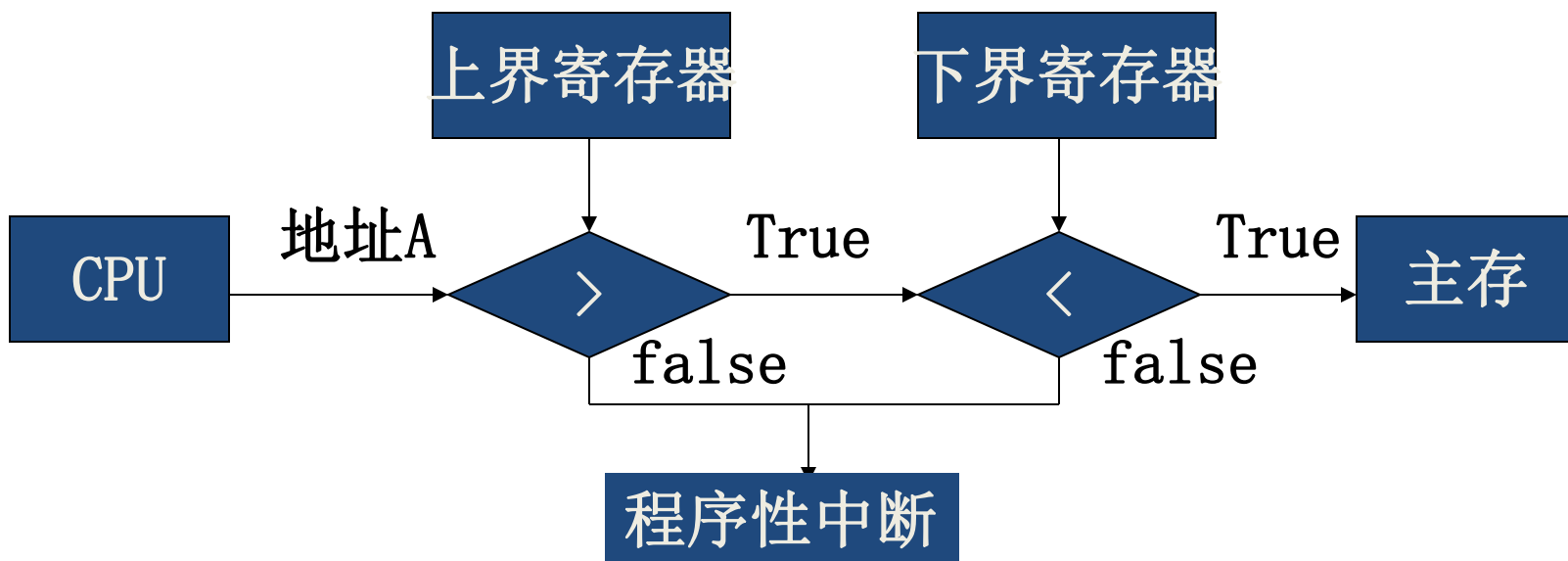
将用户内存
空间分成长度
固定的若干块。

用户空间

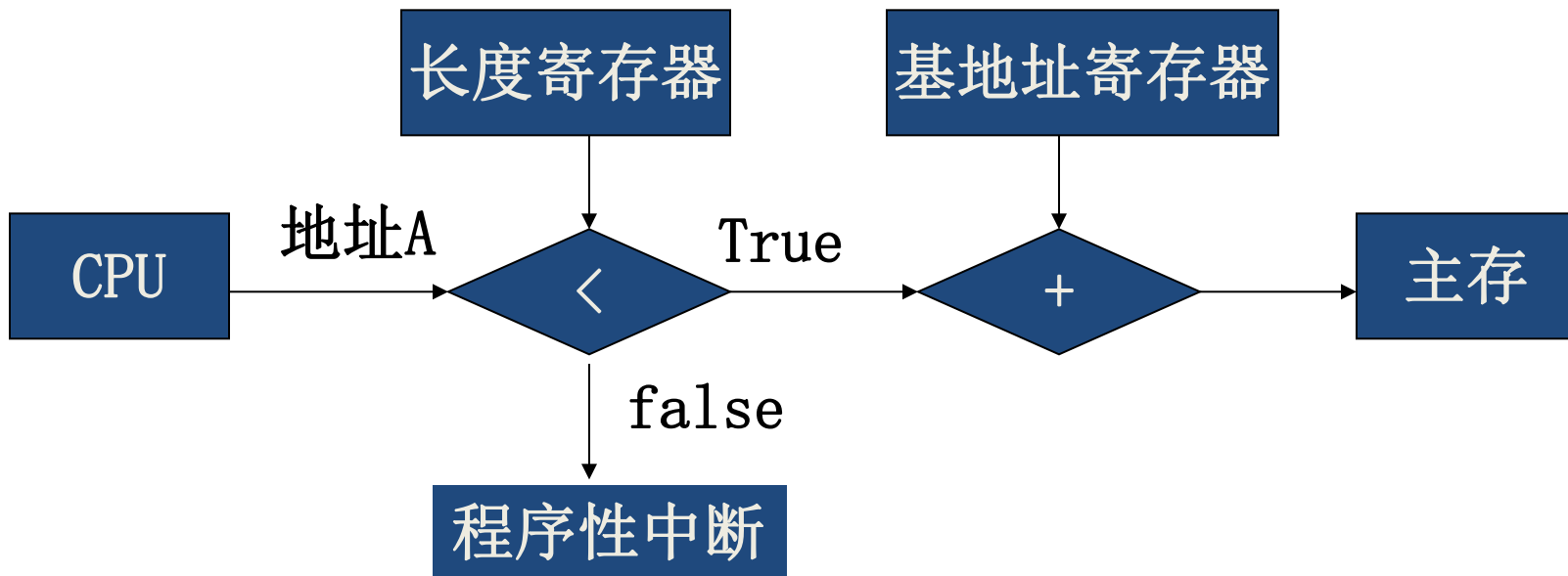


地址访问保护有两种方式：

1. **上下界寄存器和地址检查机构。**当作业被调度运行时，作业在内存中的上下界地址送上下界寄存器，每次内存访问时，地址检查机构作越界检查。作业程序须是绝对地址或静态可浮动的。



2. 基址寄存器、长度寄存器和动态地址转换机构。当作业被调度运行时，将作业所占内存基址及长度送基址、长度寄存器，每次内存访问时，先看访问地址是否小于长度，然后+基址进行访存。用户程序代码是动态浮动的。

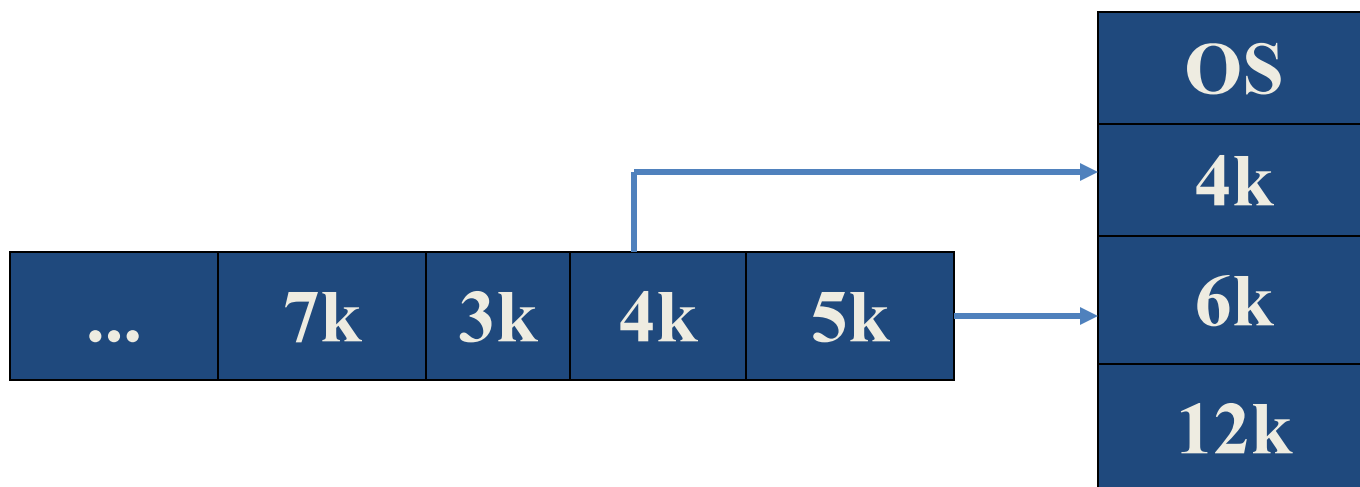


二、作业调度

多队列法



单队列法

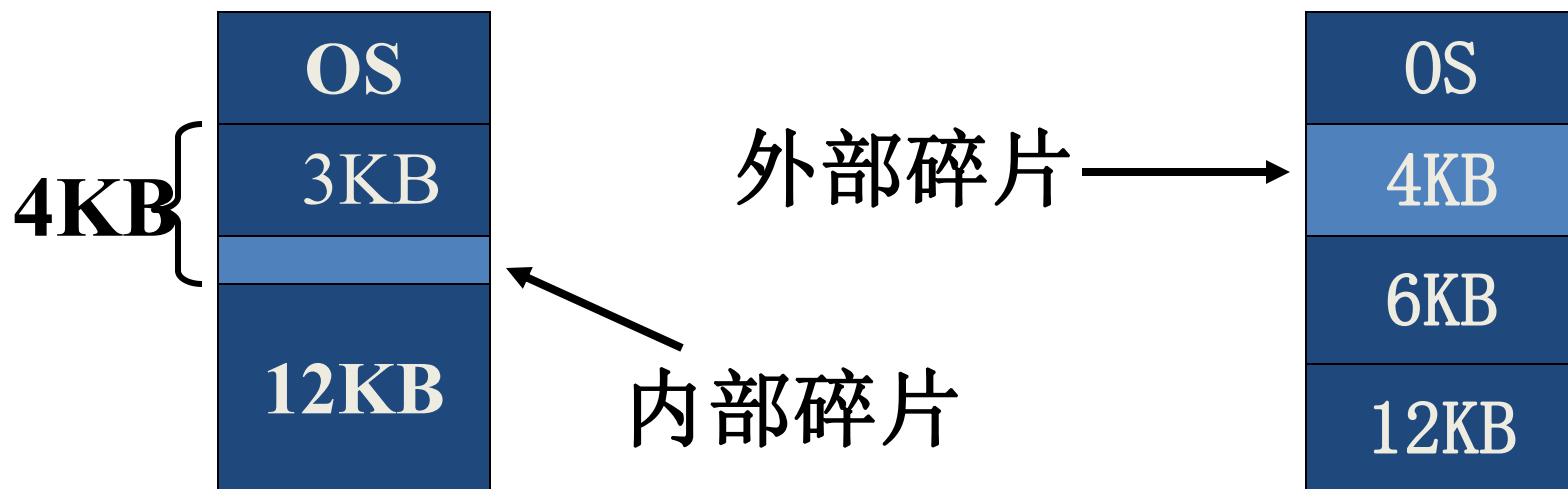


三、存储碎片

内部碎片： 内存某存储区间大于其存放作业空间的部分。

外部碎片： 内存某存储区间容不下要运行的作业时。

作业长度：5KB，8KB，12KB



4.2.3 多道连续可变分区法

特点：多道、连续、但不固定划分内存。

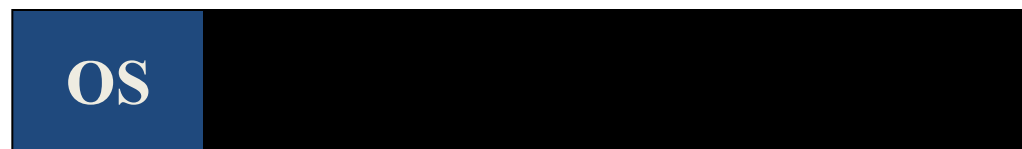
一、管理方法

系统设置一个空闲块队列，初始状态时队列中只有一个连续的空闲块。作业到达后，以某种策略分配空间。作业撤离时，将释放的空间加入空闲队列。

举例：假设任一时间段内，内存中每一作业的运行时间相等。

作业到来次序	所需存储量	运行时间
1	60KB	10s
2	100KB	5s
3	30KB	20s
4	70KB	8s
5	50KB	15s

0 40 256



分配：分配策略包括首次满足法/最佳满足法/最大满足法，在找到合适的空闲块后，从其中将作业大小的空间分给作业，而剩余部分挂入空闲队列。

下面 F 是空闲块集合； $\text{size}(k)$ 为块 k 的大小； $\text{size}(v)$ 为用户所需空间。

1. 如果所有属于 F 的 k ，均有 $\text{size}(k) < \text{size}(v)$ ，则失败。

2. 否则按某一策略选出 k ，使得

$$\text{size}(k) > \text{size}(v)$$

3. $F = F - \{k\}$ ；

(续分配)

4. 如果 $\text{size}(k) - \text{size}(v) < \text{基本单位}$, 则将 k 分给用户。

5. 否则将 k 分成 $k1, k2$, 其中

$$\text{size}(k1) = \text{size}(v), F = F + \{k2\}$$

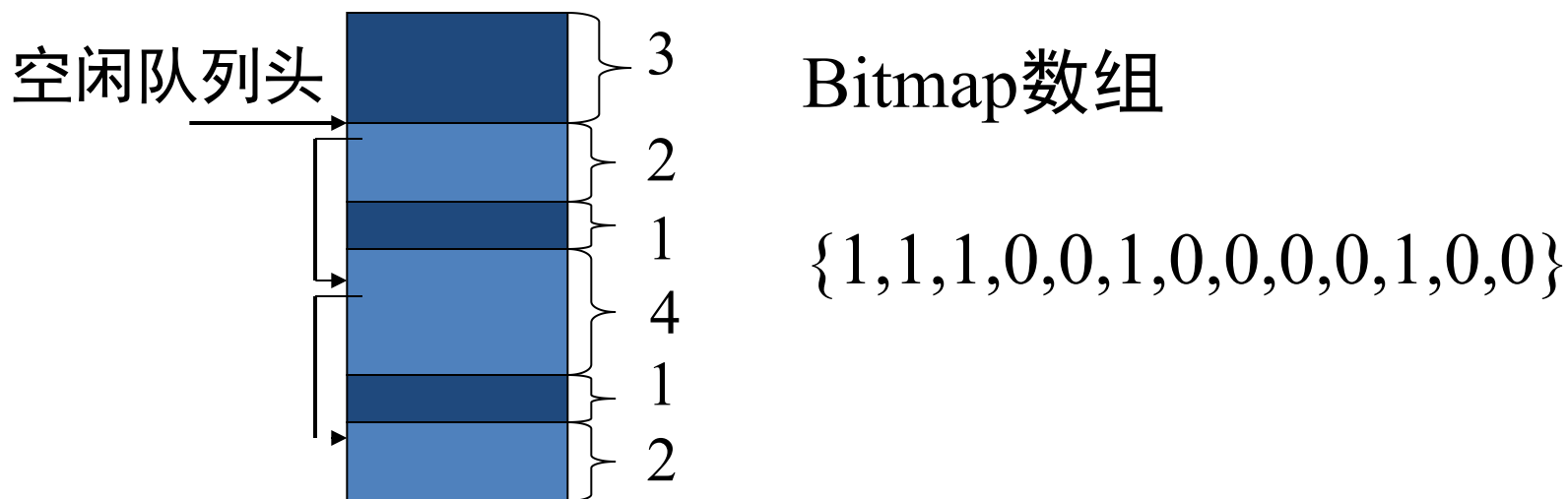
回收:

当作业结束时, 收回作业所占空间, 将此块链入空闲队列。

若空闲队列中原来有与此块的相邻块, 则把这些块合并成一个大连续块。

二、可用空间管理

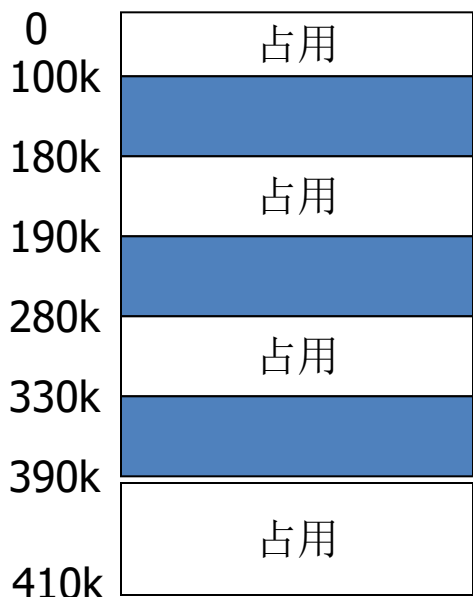
除用队列表示可用空闲块外,也可以用数组登记可用空闲块, 数组项=用户空间总量/基本分配单位。



紧致：通过移动作业位置可以将零散的空闲块连接成大块。要求作业动态可浮动。

练习一

- 设内存的分配情况如图所示。若要申请一块**40K**字节的内存空间，若采用最佳适应算法，则所得到的分区首址为_____。
- **A.100K B.190K C.330K D.410K**



答案： C

练习二

- 分区管理中采用“最佳适应”分配算法时，宜把空闲区按_____次序登记在空闲区表中。
- A. 长度递增 B. 长度递减
- C. 地址递增 D. 地址递减

答案：A

练习三

- 在固定分区分配中，每个分区的大小是()。
- A.相同 B.随作业长度变化
- C.可以不同但预先固定
- D.可以不同但根据作业长度固定

答案： C

练习四

- 首次适应算法的空闲区是_____。
- A . 按地址递增顺序连在一起
- B . 始端指针表指向最大空闲区
- C . 按大小递增顺序连在一起
- D . 寻找从最大空闲区开始

答案：A

练习五

- 把作业地址空间中使用的逻辑地址变为内存中的物理地址的过程称为_____。
- A 重定位 B 物理化
- C 逻辑化 D 加载

答案： A