



操作系统

Operating System

- 
-
- 教师：李 源
 - 电话：13708862006
 - 邮箱：liyuan@ynu.edu.cn
 - Q Q：2583695240



课程概况

- **课程特点：概念多、原理性强、较抽象**
- **课程学习目的：基础核心课、有利于对计算机系统的理解和软件开发**
- **课程学习方法：以问题驱动学习、理论联系实际**
- **课程学习难点：概念、原理、算法、数据结构**



参考资料

Operating System Internals and Design principles, William Stallings , Publishing House of Electronics Industry

- **Applied Operating System Concepts , Abraham Silberschatz *etc.* , Higher Education Press**
- **操作系统实用教程, 任爱华 *etc.* 清华大学出版社 (第三版)**



课程主要内容

- 👉 操作系统引论（第1章）
- 👉 进程管理（第2-3章）
- 👉 存储器管理（第4-5章）
- 👉 设备管理（第6章）
- 👉 文件管理（第7-8章）
- 👉 操作系统接口（第9章）



第1章 操作系统引论

- ❖ 操作系统的目标 and 作用
- ❖ 操作系统的发展过程
- ❖ 操作系统的基本特征
- ❖ 操作系统的主要功能
- ❖ 操作系统的结构设计



1.1 操作系统的目标和作用

一、计算机的系统组成

二、什么是操作系统（4种基本观点）

三、操作系统的目标

四、操作系统的作用

五、推动操作系统发展的主要动力

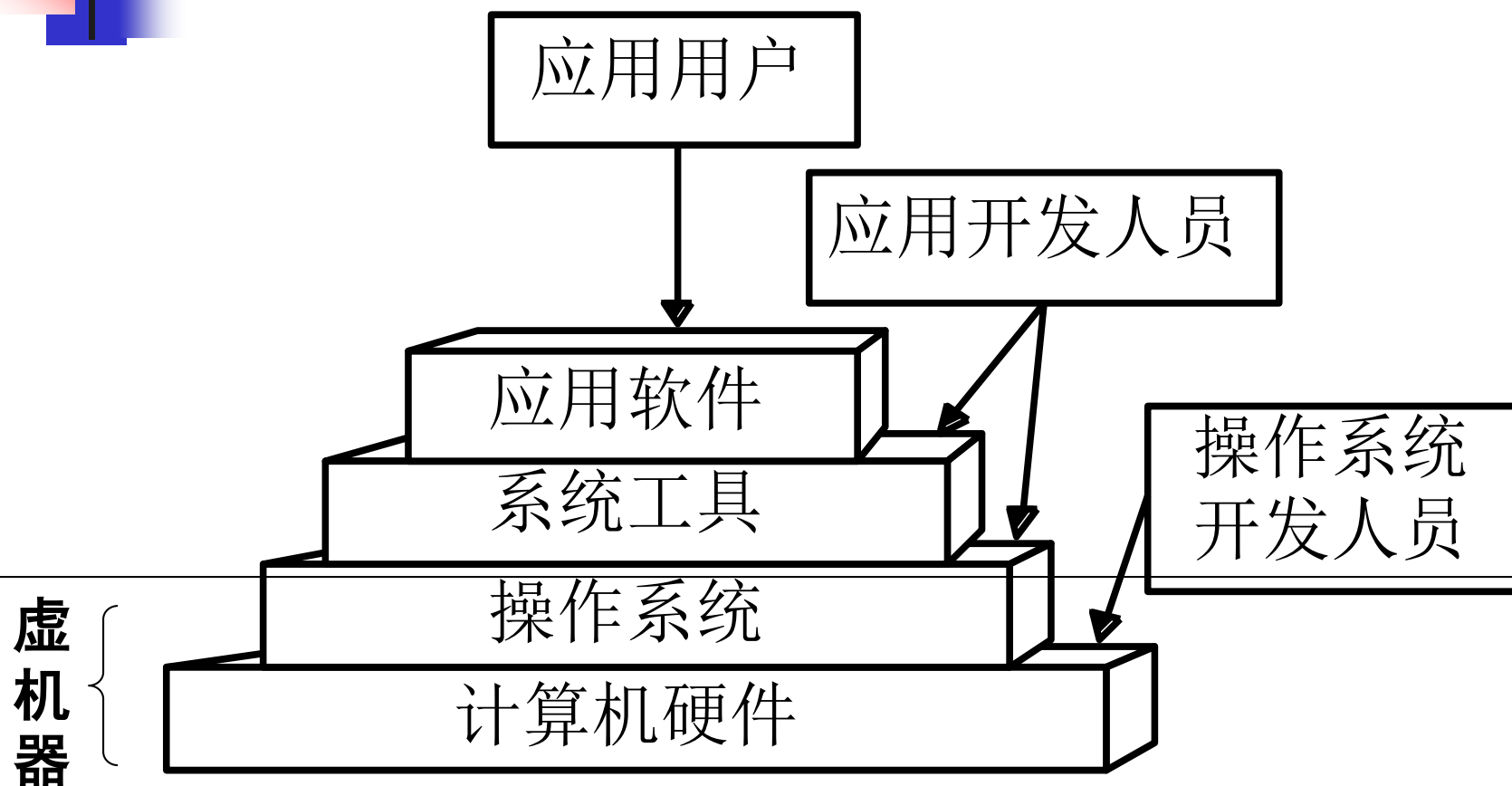


一、计算机的系统组成

1、计算机系统的资源分类

- **软件：系统软件（操作系统及实用程序）、应用软件**
- **硬件：处理机（CPU）、内存（条）、外部I/O设备（显示器、键盘、鼠标、打印机、磁盘、磁带、扫描仪...），以及系统总线**

2. 计算机系统的层次结构



操作系统的地位：紧贴系统硬件之上，所有其他软件之下（是其他软件的共同环境）



二、什么是操作系统

关于现代操作系统的4种基本观点

➤ 从外部看操作系统

- 计算机用户的观点：用户环境观点
- 应用程序员的观点：虚拟机器观点

➤ 从内部看操作系统

- OS开发者观点之一：资源管理观点
- OS开发者观点之二：作业组织观点



用户环境观点

- 该观点认为，操作系统是计算机用户使用计算机系统的接口，它为计算机用户提供了方便的工作环境。
- 计算机用户：终端用户、程序员和系统设计者。
- 操作系统提供的接口
 - 用户接口（User Interface，也称为命令接口）
 - 程序接口（系统调用，也称为应用编程接口，Application Programming Interface，简称API）



虚拟机器观点

- **该观点认为，操作系统是建立在计算机硬件平台上的虚拟机器，它为应用软件提供了许多比计算机硬件功能更强或计算机硬件所没有的功能。**
- **操作系统在虚拟机中充当管理员和协调员的角色，管理计算机的软硬件资源，并协调多任务、多进程的运行。**
- **扩充：功能、计算机数量**



资源管理观点

- **该观点认为，操作系统是计算机系统中各类资源的管理者，它负责分配、回收以及控制系统中的各种软硬件资源。**
- **跟踪资源的使用状况、满足资源请求、提高资源利用率,以及协调各程序和用户对资源的使用冲突。**
 - **监视资源**
 - **分配/回收资源**
 - **保护资源**



作业组织观点

- **该观点认为，操作系统是计算机系统工作流程的组织者，它负责协调在系统中运行的各个应用程序的运行次序。**
- **用于巨型机和大型服务器上，以批文件方式提交作业，请求主机逐个运行。**
- **主机操作系统负责组织、协调各个作业的运行，报告执行结果或错误信息。**
- **减少了人工干预，提高了系统的效率。这种工作方式有利于有效利用造价高且性能强大的主机资源。**



操作系统的定义

操作系统是计算机系统中的一个系统软件，用于管理和控制计算机系统中的硬件和软件资源，合理地组织计算机的工作流程，以便有效利用这些资源，为用户提供一个功能强、使用方便的工作环境，从而在计算机与用户之间起到接口的作用。



三、操作系统的目标

- **有效性**：以一种有效的方式使用计算机的资源
- **方便性**：使计算机系统更易于被用户使用
- **可扩充性**：采用新的OS结构, 易于功能的增、删、改
- **开放性**：要求统一开放的环境，各种类型的计算机硬件系统，出自不同的厂家，能通过网络集成并能够正确、有效地协同工作，实现应用程序的移植。



四、操作系统的作用（1）

- OS作为用户与计算机硬件系统之间的接口
 - 命令方式
 - 系统调用方式
 - 图形、窗口方式



四、操作系统的作用（2）

- OS作为计算机系统资源的管理者
 - 管理对象包括：CPU、存储器、外部设备、信息（数据和软件）；
- 管理的内容：资源的当前状态（数量和使用情况）、资源的分配、回收和访问操作，相应管理策略（包括用户权限）



四、操作系统的作用（3）

➤ OS实现了对计算机资源的抽象

- 在裸机上添加：设备管理、文件管理、存储管理（针对内存和外存）、处理机管理（针对CPU）；
- 即OS用作扩充机器（extended machine）/虚拟机（virtual machine）。



五、推动操作系统发展的主要动力

- ❑ 不断提高计算机资源利用率的需要
- ❑ 方便用户
- ❑ 器件的不断更新换代
- ❑ 计算机体系结构的不断发展



1.2 操作系统的发展过程

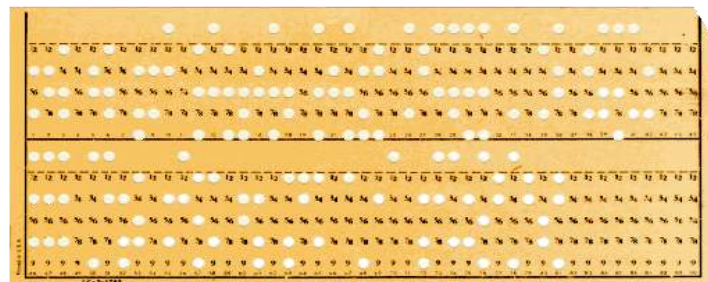
- 无操作系统的计算机系统
 - 人工操作方式
 - 脱机输入输出方式
- 批处理系统
 - 单道批处理系统
 - 多道批处理系统
- 分时系统
- 实时系统
- 微机操作系统
- 多处理机操作系统
- 网络操作系统
- 分布式操作系统
- 嵌入式操作系统
- 智能卡操作系统

1.2.1 无操作系统的计算机系统

(1) 人工操作方式

■ 工作流程：

- (1) 由程序员事先穿孔（对应程序和数据）
- (2) 将穿孔的纸带（卡片）装入纸带（卡片）输入机
- (3) 再启动输入机将程序和数据输入计算机，
- (4) 然后启动计算机运行
- (5) 运行完毕取走计算机结果
- (6) 下一位用户





人工操作方式带来的矛盾及解决方法

- **缺点：**

- (1) 用户独占全机（资源浪费）

- (2) CPU等待人工操作：装入、取走时CPU及内存等资源空闲。CPU既负责计算，又要负责传输。（属于单控制方式）

- **矛盾**

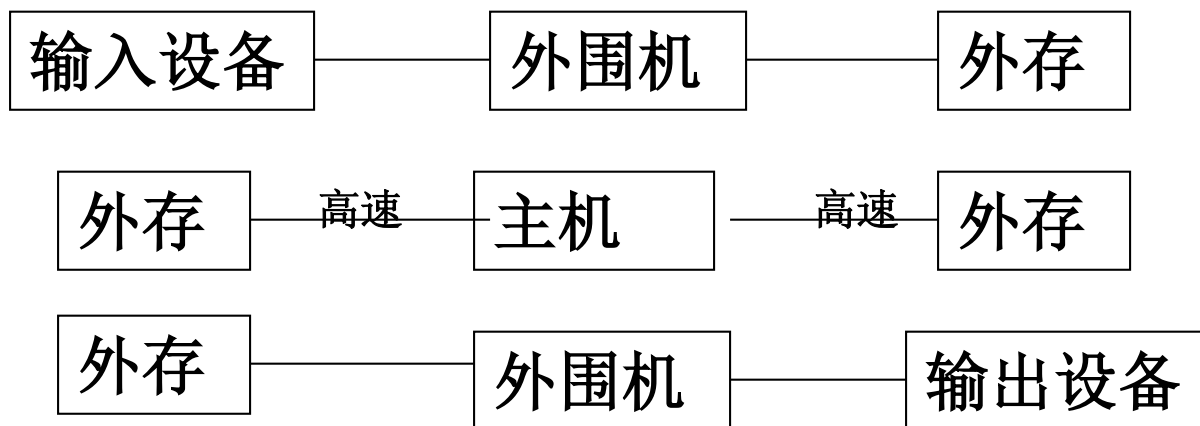
- 人机矛盾即人工操作方式与机器利用率的矛盾
 - CPU与I/O设备之间速度不匹配的矛盾

- **解决方法**

- 脱机输入输出方式

脱机输入/输出方式 (off-Line I/O)

■ 工作流程：（示意图）

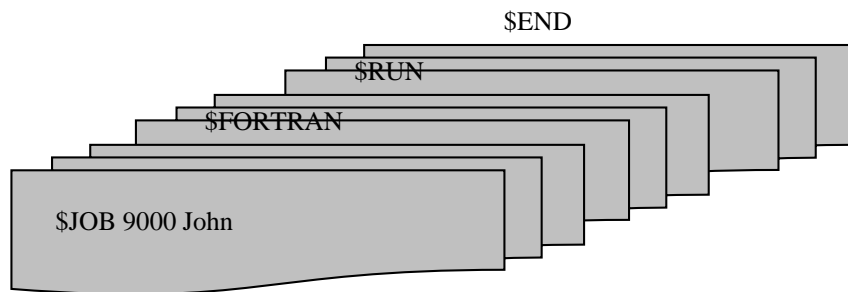


■ 优点：

- (1) 减少了CPU的空闲时间。
- (2) 提高了I/O速度（与高速的磁盘打交道）

1.2.2 单道批处理系统

- 程序员首先将命令、程序和数据用汇编语言或 FORTRAN 语言写在纸上,然后用穿孔机制成卡片,最后将这些卡片交给操作员



- 批处理程序(又称为监督程序,或管理程序),管理应用程序的运行。



单道批处理系统的操作步骤

操作步骤:

- 1.收集一批作业(卡),用专用I/O计算机将作业逐个读到磁带上保存起来。**
- 2.批处理程序将磁带上的第一个作业读入计算机,运算结束后将结果输出到输出磁带上。**
- 3.自动读入下一个作业,并运行。**
- 4.当一批作业全部执行结束后,取下输入磁带和输出磁带,用输入磁带录入下一批作业,将输出磁带送到专用输出计算机,进行脱机打印。**



单道批处理系统的基本特性

- 自动性：一批作业自动逐个依次执行。
- 顺序性：作业顺序进入内存，顺序执行。
- 单道性：内存中只有一道程序运行。



单道批处理系统的评价

- 解决了作业间的自动转接问题,减少了机器时间的浪费。
- 不管作业大小,只要它一旦占用处理机开始执行,则它必须一直占据处理机,直到运行完毕。
- 资源利用率低。
- 对短作业不公平,因为它们等待执行的时间可能远远超过它们实际执行的时间。
- 交互性差。作业由批处理程序控制运行,用户无法实时控制,如果运行中途出现故障,也只能停下来,重新运行。



1. 2. 3 多道批处理系统

- **单道批处理系统中，任意时刻只允许一道作业在内存中运行，资源利用率低。**
- **为了提高系统资源利用率和系统吞吐量，形成了多道批处理系统。**



多道批处理系统

- **多道**是指，允许多个程序同时存在于主存中，按照某种原则分派处理机，逐个执行这些程序。
- **批处理**：用户提交的作业首先存放在外存，并排成一个队列。然后，由作业调度程序按照一定的算法从该队列中一次选取一个或若干个作业装入内存执行。

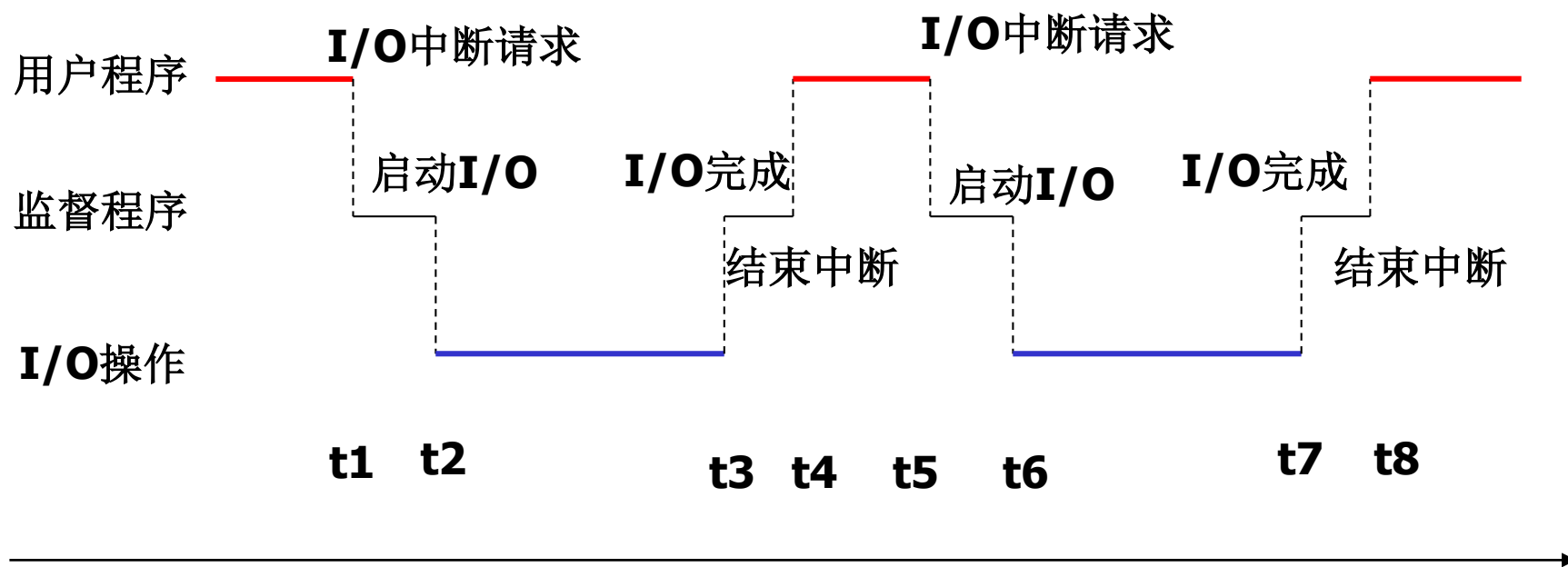


多道批处理系统

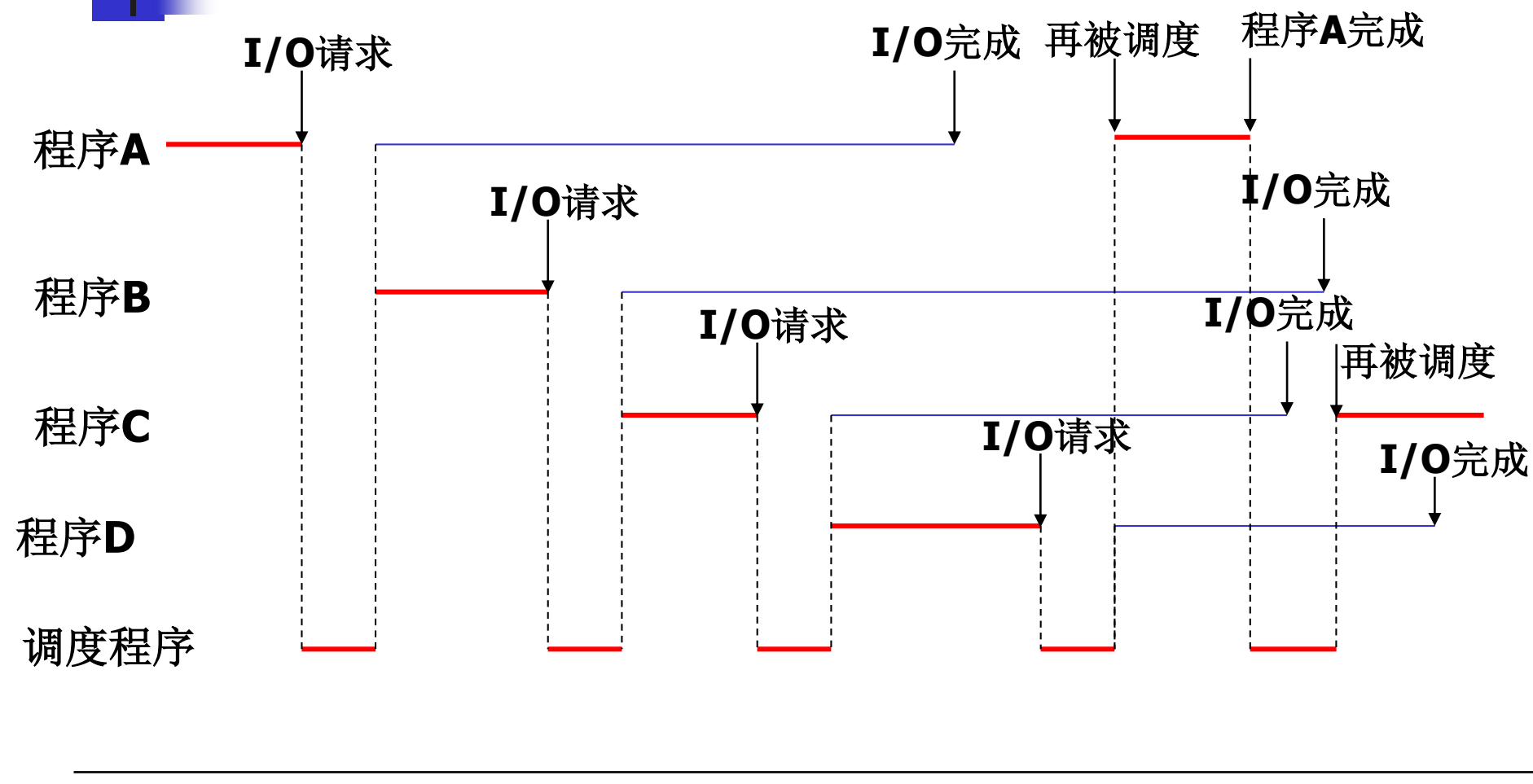
处理机自动切换

- ▶ **当某个程序占用处理机执行过程中遇到了输入/输出语句，可以启动专门负责输入/输出的系统服务程序完成输入/输出操作，而处理机切换到另一个程序执行**

单道程序运行情况



四道程序运行情况





多道程序设计技术

- 为了提高系统吞吐量和资源利用率，允许多个程序同时驻留内存，使处理机在这些程序之间切换，在一段时间内，执行完多个程序的处理技术称为多道程序设计技术（multiprogramming）。
- 现代操作系统大多都采用了多道程序设计技术。



多道程序设计技术引发的问题

- **处理机的分配与回收**
- **内存的分配与保护**
- **I/O设备的共享与效率**
- **文件的有效管理**
- **作业的组织**



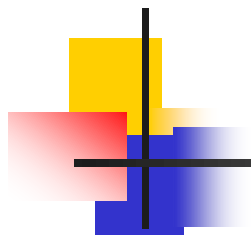
多道批处理系统的优缺点

- 优点：

- 资源利用率高
- 系统吞吐量大

- 缺点：

- 平均周转时间长
- 无交互能力

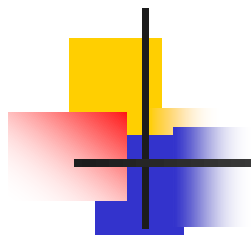


例1 有两个程序A和B，A程序执行时所作的工作按次序需要的时间为：CPU计10s，设备1计5s，CPU计5s，设备2计10s，CPU计10s。B程序执行时所作的工作按次序需要的时间为：设备1计10s，CPU计10s，设备2计5s，CPU计5s，设备2计10s。问在单道方式下和多道并发环境下执行A、B两个程序，CPU的利用率分别是多少？
(调度程序的执行时间忽略不计)

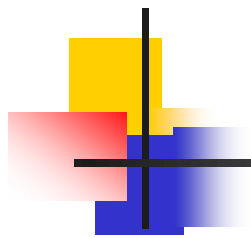


甘特图 (Gantt chart)

- 甘特图 (Gantt chart) 又叫横道图。它是在第一次世界大战时期发明的，以亨利·劳伦斯·甘特先生的名字命名，他制定了一个完整地用条形图表进度的标志系统。
- 甘特图内在思想简单，即以图示的方式通过活动列表和时间刻度形象地表示出任何特定项目的活动顺序与持续时间。基本是一条线条图，横轴表示时间，纵轴表示活动（项目），线条表示在整个期间上计划和实际的活动完成情况。它直观地表明任务计划在什么时候进行，及实际进展与计划要求的对比等。
- 应用领域：通信科技；政策法规、管理科学



例1 有两个程序A和B，A程序执行时所作的工作按次序需要的时间为：CPU计10s，设备1计5s，CPU计5s，设备2计10s，CPU计10s。B程序执行时所作的工作按次序需要的时间为：设备1计10s，CPU计10s，设备2计5s，CPU计5s，设备2计10s。问在单道方式下和多道并发环境下执行A、B两个程序，CPU的利用率分别是多少？
（调度程序的执行时间忽略不计）



例2 一个多道批处理系统中仅有P1和P2两个作业，P2比P1晚5ms到达，它们的计算和I/O操作顺序如下：

P1：计算 60ms，I/O 80ms，计算 20ms；

P2：计算 120ms，I/O 40ms，计算 40ms。

若不考虑调度和切换时间，求完成两个作业需要的最少时间。

练习

设内存中有三道程序A、B、C，它们按A、B、C的优先次序执行。它们的计算和I/O操作时间如下表：（单位：ms）

程序 操作	A	B	C
计算	30	60	20
I/O	40	30	40
计算	10	10	20

假设三道程序使用相同的设备进行I/O操作，即程序以串行方式使用设备，试画出单道运行和多道运行的时间关系图（调度程序的执行时间忽略不计），并由此分别给出两种方式下完成这三道程序所花的时间。

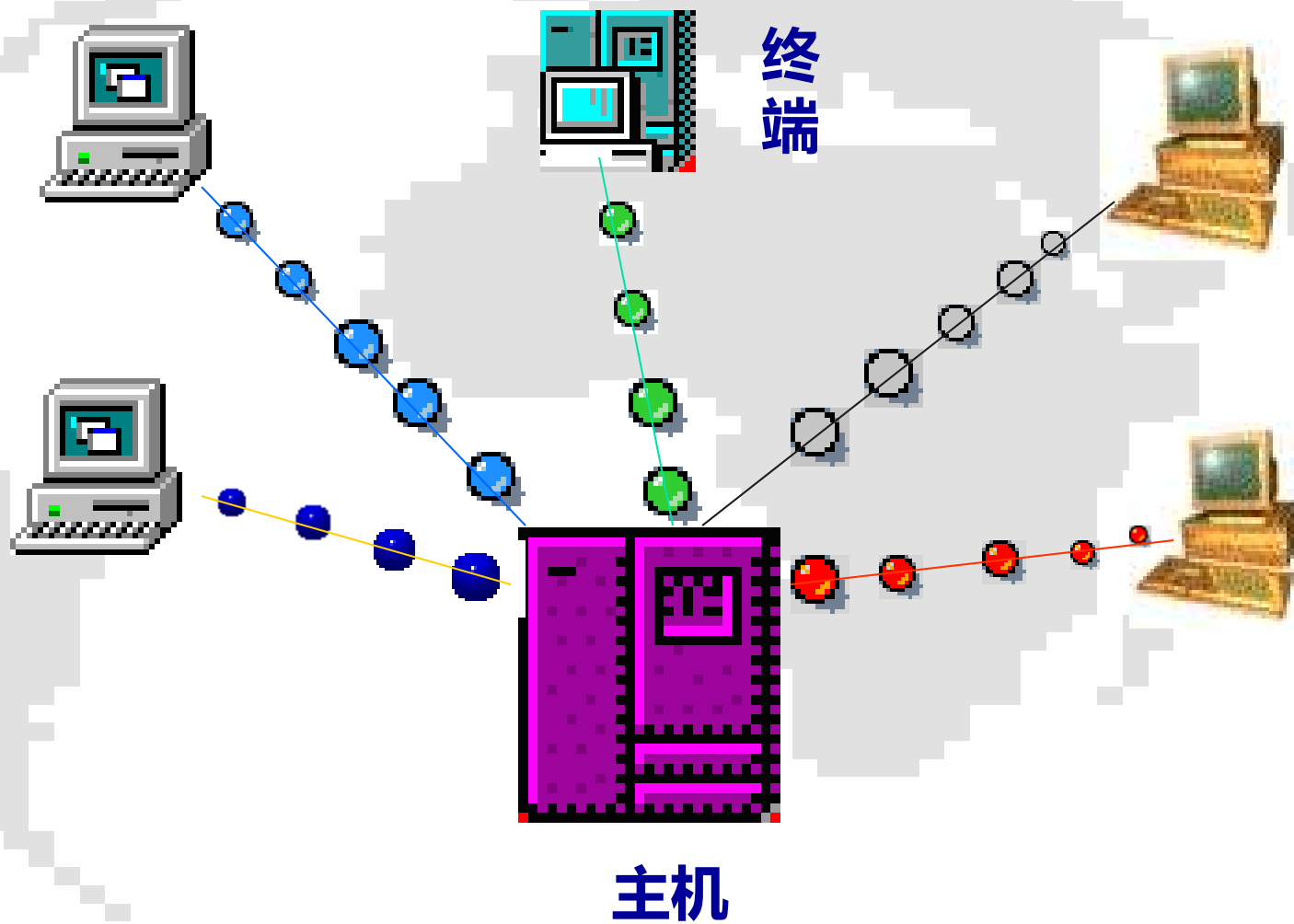


1.2.4 分时系统

■ 分时系统(Time Sharing System)的产生

人-机交互、共享主机、便于用户上机

- **分时系统**：在一台主机上连接了多个带有显示器和键盘的终端，同时允许多个用户通过自己的终端，以交互方式使用计算机，共享主机中的资源





分时系统实现中的关键问题

- **及时接收：**通过多路卡同时接收各用户输入的数据，存入缓冲区。
- **及时处理：**各用户的作业驻留内存，每个作业只运行一个时间片。



分时系统的特征

- 多路性
- 独立性
- 及时性
- 交互性



1.2.5 实时系统 (Real Time System)

实时操作系统：

指系统能及时（或即时）响应外部事件的请求，在规定的时间内完成对该事件的处理，并控制所有实时任务协调一致地运行。

■ 实时操作系统的应用领域

有实时要求（有严格确定的时间限制）的领域（过程控制及事务处理）——多道批处理系统和分时系统不能满足



实时系统 (Real Time System)

实时系统分类：

- (1) 实时控制系统
- (2) 实时信息处理系统

■ 实时任务：

周期性/非周期性实时任务（截止时间）

硬实时任务/软实时任务

■ 实时系统与分时系统特征的比较：

多路性、独立性、及时性、交互性、可靠性



三种基本的操作系统

■ 三种基本的操作系统

批处理操作系统、分时操作系统和实时操作系统

■ 通用操作系统

如果一个操作系统兼有批处理、分时和实时系统三者或二者的功能，则称该操作系统为通用操作系统。



1.2.6 微机操作系统

■ 微机操作系统的特征

- **应用领域**：事务处理、个人娱乐，
- **系统要求**：使用方便、支持多种硬件和外部设备（多媒体设备、网络、远程通信）、效率不必很高。

■ 常用的微机操作系统

- **单用户单任务**：CP/M, MS-DOS
- **单用户多任务**：Windows
- **多用户多任务**：Solaris OS, Linux。



1.2.7 多处理机操作系统

■ 多处理机系统引入原因

- 增加系统的吞吐量
- 规模经济、节省投资（共享外设、存储器等）
- 提高系统的可靠性

■ 多处理机系统的类型

- 紧密耦合MPS
- 松散耦合MPS

■ 多处理机操作系统的类型

- 非对称多处理模式（每个处理器都有各自特定的任务）
- 对称多处理模式（每个处理器都运行同一个操作系统的拷贝）



1.2.7 网络操作系统

- 计算机网络的类型（按拓扑结构分为：星型、树型、总线型、环型、网状型）
- 网络OS的模式
 - 客户/服务器模式（C/S）
 - 对等模式（P2P）
- 网络OS的功能
 - 通信、资源管理、网络服务、网络管理、互操作能力



1.2.8 分布式操作系统

- **分布式系统**：能将任务**处理和控制**进行分散/分布的系统（相对于集中式）。
- **分布式系统的两种环境**：多处理机系统；多计算机系统。
- **分布式OS与网络OS的比较**
 - 1. 分布性 2. 并行性 3. 透明性
 - 4. 共享性 5. 健壮性

分布式系统基本特征——处理上的分布，即功能和任务的分布。



1.3 操作系统的基本特征

- 并发 (concurrency)
- 共享 (sharing)
- 虚拟 (virtual)
- 异步性 (asynchronism)



并发 (concurrency) (1)

■ **并发**：指在计算机系统中同时存在着多道运行的程序（**进程**）

宏观上：多道程序**同时**在执行

微观上：任何时刻**只有一道**程序在执行，
即微观上多道程序在CPU上**轮流（交替）**执行
（单机）

■ **并行** (parallel)：与并发相似，指多道程序在同一时刻执行，但需多个硬件支持。



并发 (concurrency) (2)

■ 引入进程

程序和进程：在多道程序系统中，程序不能独立运行，更不能和其它程序并发执行。引入进程的目的，就是为了使多个程序并发执行。

■ 引入线程

进程是OS中拥有资源并独立运行的基本单位。引入线程后，把进程作为分配资源的基本单位，把线程作为独立运行和独立调度的基本单位，进一步提高系统的并发性。



并发 (concurrency) (3)

注：1、程序的并发执行能有效改善系统资源的利用率，但使系统复杂化，因此操作系统必须对并发活动进行控制和管理。

2、并发是操作系统最重要的特征，其它特征均以并发为前提。



共享 (sharing)

■指系统中的资源不再为某道程序所独占，而是供多道程序共同使用。

■资源共享方式：互斥共享方式、同时访问方式

■并发和共享是操作系统的两个最基本的特征，二者互为存在条件：

- 资源的共享是以程序（进程）的并发执行为条件。
- 程序的并发执行也以资源的共享为条件。



虚拟(virtual)

- 通过某种技术把一个**物理实体**映射为若干个对应的**逻辑实体**——分时（时分复用技术）或分空间（空分复用技术）。
- 虚拟是操作系统管理系统资源的重要手段，可提高资源利用率。如：虚拟处理机（分时）、虚拟设备（分时）、虚拟磁盘（分空间）、虚拟存储器（分空间、分时）



异步性 (asynchronism)

- 也称不确定性，指进程在执行中，其执行时间、顺序、向前推进的速度和完成的时间等都是不可预知的。
- 只要在OS中配置有完善的进程同步机制，且运行环境相同，作业经多次运行都会获得完全相同的结果。所以，异步运行方式是允许的，是OS的一个重要特征。



1.4 操作系统的功能

- 处理机管理
- 存储器管理
- 设备管理
- 文件管理
- 用户接口



一、处理机管理的功能

- 处理机管理的主要任务：

- 是对处理机进行**分配**
- 对处理机运行**进行有效的控制和管理**

注：处理机的分配和运行以进程为基本单位，因此
对处理机的管理可归结为对进程的管理

- 处理机管理的功能

- 进程控制
- 进程同步
- 进程通信
- 调度



二、存储器管理的功能

■ 主要任务

- 为多道程序的运行提供良好的环境
- 方便用户使用存储器
- 提高存储器的利用率
- 从逻辑上扩充内存

■ 功能

- 内存分配（静态分配、动态分配）
- 内存保护
- 地址映射
- 内存扩充



三、设备管理的功能

■ 主要任务

- 完成用户提出的I/O请求
- 为用户分配I/O设备
- 提高I/O设备的利用率及速度
- 方便用户使用I/O设备

■ 功能

- 缓冲管理
- 设备分配
- 设备处理
- 虚拟设备



四、文件管理的功能

- 主要任务

- 对用户文件和系统文件进行管理
- 方便用户使用文件
- 保证文件的安全性

- 功能

- 文件存储空间的管理
- 目录管理
- 文件的读、写管理和保护



五、OS与用户接口管理的功能

- 主要任务

- 方便用户使用操作系统

- 功能

- 用户接口（联机用户接口-命令方式、图形用户接口，脱机用户接口）
 - 程序接口（系统调用）



1.5 操作系统的结构设计

操作系统的结构设计经历了以下几代：

- 传统的操作系统结构
 - 无结构操作系统
 - 模块化结构OS
 - 分层式结构OS
- 客户/服务器模式
- 面向对象的程序设计
- 微内核OS结构



无结构操作系统

- OS是由众多的过程直接构成，各过程之间可相互调用，但OS内部不存在任何结构，所以这种OS是无结构的，又称为整体系统结构。

- 缺点：

既庞大又杂乱，缺乏清晰的程序结构；程序错误多，调试难、阅读难、理解难、维护难。



模块化结构操作系统（1）

- OS是采用“模块化程序设计”技术，按其功能划分为若干个独立的模块，管理相应的功能，同时规定好各模块之间的接口，以实现它们的交互，对较大模块又可按子功能进一步细分下去。所以这种OS称为模块化OS结构。
- 模块的独立性
 - 关键：模块的划分和规定好模块之间的接口
 - 衡量独立性的两个标准：
 - 内聚性：指模块内部各部分间联系的紧密程度
 - 耦合度：指模块间相互联系和相互影响的程度



模块化结构操作系统（2）

■ 优点

- 提高了OS设计的正确性、可理解性和可维护性
- 增强了OS的可适用性
- 加速了OS的开发过程

■ 缺点：

- 模块及接口划分较困难
- 未区别共享资源和独占资源
- 由于管理上的差异，使OS结构变得不够清晰



分层式结构操作系统（1）

- 分层式结构是对模块化结构的一种改进，它按分层式结构设计的基本原则，将OS 划分为若干个层次，每一层都只能使用其底层所提供的功能和服务，从硬件开始，在其上面一层一层地自底向上增添相应功能的软件，这种OS结构称为分层式OS结构。
- 特点：
 - 每一步设计都建立在可靠的基础上，结构更清晰
 - 调试和验证更容易，正确性更高



分层式结构操作系统（2）

- 分层式结构的优点：
 - 易保证系统的正确性
 - 易扩充和易维护
- 缺点：
 - 系统效率降低：层次结构是分层单向依赖的，必须在相邻层次之间建立层次间的通信机制，增加了系统开销。



客户/服务器模式（1）

■ 客户/服务器模式的组成

- 客户机：每台客户机都是一个自主计算机，客户进程在运行
- 服务器：一台规模较大的机器，为网上所有的用户提供一种或多种服务。
- 网络系统：用于连接所有客户机和服务器，实现它们之间的通信和网络资源共享。

■ 客户/服务器之间的交互：

- 客户发送请求消息、服务器接收消息
- 服务器回送消息、客户机接收消息



客户/服务器模式（2）

- 客户/服务器模式的优点

- 数据的分布处理和存储。
- 便于集中管理。
- 灵活性和可靠性。
- 易于改编应用软件

- 不足之处

- 存在着不可靠性和瓶颈问题
- 改进方法：在网络中配置多个服务器，并采取相应的安全措施。



面向对象的程序设计

- 面向对象技术的基本概念
 - 对象
 - 类
 - 继承
- 面向对象技术的优点
 - 通过“重用”提高产品质量和生产率。
 - 使系统具有更好的易修改性和易扩展性。
 - 更易于保证“正确性”和“可靠性”



微内核的OS结构（1）

- 微内核技术的主要思想

在OS内核中只留下一些最基本的功能，而将其他服务分离出去，由工作在用户态下的进程来实现，形成“客户/服务器”模式。客户进程可通过内核向服务器进程发送请求，以取得OS的服务。

- 微内核

精心设计的，能实现现代OS核心功能的小型内核，它小而精炼，运行在核心态下，开机后常驻内存，不会因内存紧张而换出，它为构建通用OS提供了一个重要基础。



微内核的OS结构（2）

- 微内核OS的基本概念

在进行现代OS结构设计时，大多采用基于客户/服务器模式的微内核结构，将OS分为两部分：微内核和多个服务器。

- 足够小的内核
- 基于客户/服务器模式
- 应用“机制与策略分离”原理
- 采用面向对象技术



微内核的OS结构（3）

- 微内核的基本功能
 - 进程（线程）管理
 - 低级存储器管理
 - 中断和陷入处理
- 优点：微内核OS结构是建立在模块化、层次化结构的基础上的，并采用了C/S模式和OO的程序设计技术
 - 提高了系统的可扩展性
 - 增强了系统的可靠性
 - 提供了对分布式系统的支持
 - 融入了面向对象技术



微内核的OS结构（4）

- 微内核OS存在的问题
 - 与早期的OS相比，微内核OS的运行效率有所降低
 - 最主要的原因：P32



本章要求：

了解操作系统的基本概念、操作系统发展历史，操作系统的特征、操作系统应该具备的功能、OS结构设计，掌握多道批处理系统、实时系统和分时系统的工作原理，实时和分时系统的区别。