

计算机系统的组成

1. 计算机系统是由硬件和软件组成的，硬件是软件建立与多动的的基础；软件是对硬件进行管理和功能扩充
2. 计算机硬件结构由五大功能部件组成，即运算器、控制器、存储器、输入设备和输出设备。他们经由系统总线连接在一起，实现彼此通信。其中运算器和控制器集成在一片大规模或超大规模集成电路上，称为CPU。

计算机硬件系统

计算机体系结构基本上仍沿用冯 诺依曼系统结构

要点

1. 计算机的数制采用二进制
2. 采用存储程序方式，指令和数据不加区分混合存储在同一个存储器中
3. 计算机按照人们事前指定的计算顺序来执行

功能

1. 把需要的程序和数据送至计算机中
2. 必须具有长期记忆程序、数据、中间结果及最终运算结果的能力
3. 能够万和城呢个各种算术、逻辑运算和数据传送等数据加工处理的能力
4. 能够按照要求将处理结果输出给用户

CPU

CPU在每个工作的基本周期内从主存中提取指令，对其进行解码，然后执行指令。CPU重复取指，解码并执行下一条指令，直至程序中的所有指令执行完毕。

CPU中的寄存器

- 存储器数据寄存器 MDR
- 指令寄存器 IR
- 程序计数器 PC
- 存储器地址寄存器 MAR
- 程序状态字寄存器 PSW
- 通用寄存器 R0~R3

计算机字

也成计算机字长或处理机字长、机器字长，它是指计算机进行数据处理时，一次存取。加工和传送的数据长度。一个字通常由一个或多个（一般是字节的整数倍）字节构成。字长越长，计算机一次处理的信息位就越多，精度就越高，其性能就越优越

性能指标

- 计算机字
- 主频
- CPU执行时间
- 平均指令周期数
- 百万条指令每秒
- 每秒浮点操作数

储存器

存储器是计算机系统中的记忆设备，用来存放程序和数据。

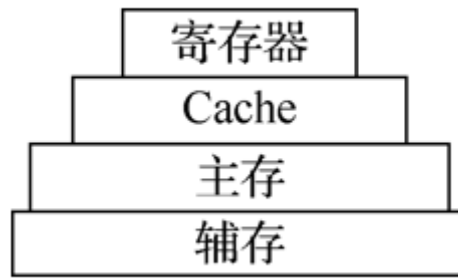
存储容量

计算机存储器通常用字节或字节组合来计算或操作

性能指标

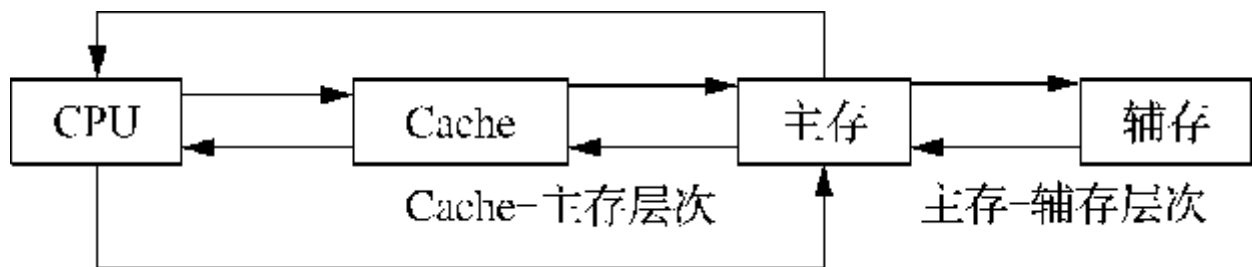
- 速度
- 容量
- 位价

一般来说，速度越快，位价就越高；容量越大，位价就越低，而且容量大，速度必然越低



层次结构

存储器的层次结构主要体现在Cache-主存和主存-辅存这两个存储层次上。



Cache能直接和CPU、主存交换信息；主存可以和CPU、Cache、外存交换信息

- Cache-主存：从CPU角度看这一层次速度接近CPU，高于主存，其容量和位价却接近主存
- 主存-辅存：从整体分析其速度接近于主存，容量接近于辅存，平均位价也接近于辅存位价，这解决了速度，容量，成本三者之间的矛盾

I/O设备

I/O设备通常由设备控制器和设备本身组成的。设备控制器是插在电路板上的一块或一组芯片，是I/O设备的电子部分，它协调和控制一台或多台I/O设备的操作，实现设备操作与整个系统操作的同步。

分类（工作方式）

- 字符设备
- 块设备
- 网络通信设备

总线

计算机系统的五大部件通过一组公共信息传输线进行连接，形成一个完整的系统，实现其功能。

分类

- 内部总线：主要连接CPU内部各个寄存器和运算部件。
- 系统总线：指CPU、主存、I/O设备各大部件之间的信息传输线，由于这些部件通常都安放在主板或各个插件板中，故又称板机总线。
- 通信总线：用于计算机系统之间或计算机系统与其他系统之间的通信。通信总线类别按照传输方式分为串行通信总线和并行通信总线。

计算机软件系统

计算机的软件系统是指在计算机中运行的各种程序，数据及相关的文档资料。

分类

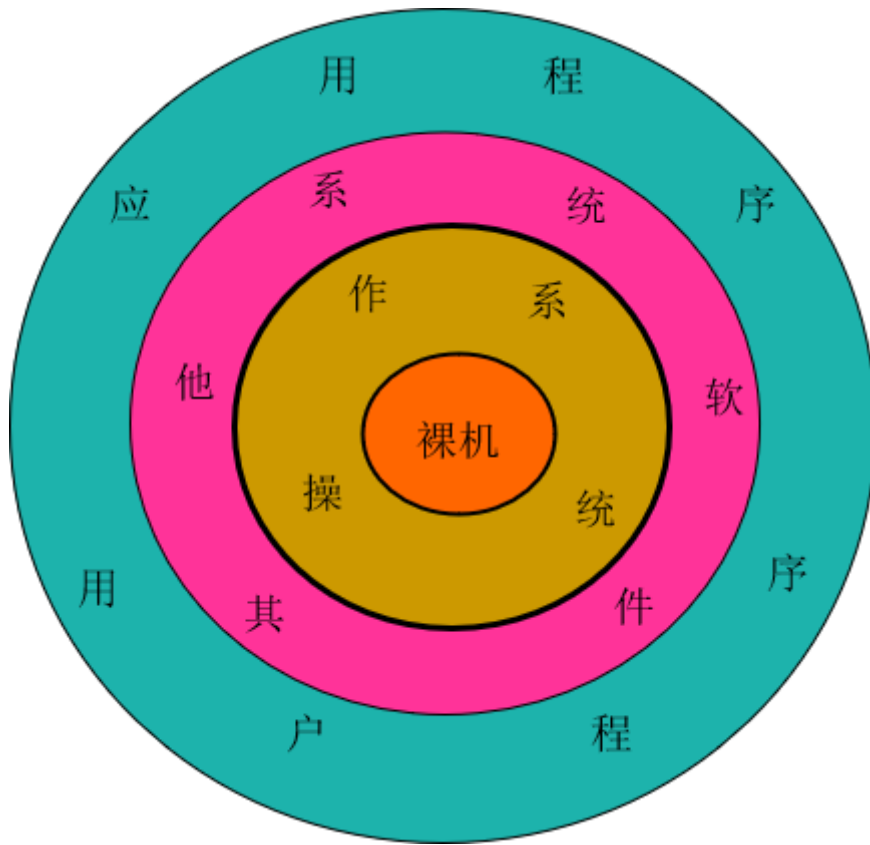
系统软件：操作系统、数据库管理系统、语言处理程序和各種服务性程序。

应用软件：办公软件、图形图像处理软件、网络通信软件、手机APP软件等。

数据结构

- 数组、线性表、堆栈和队列
- 树，树是一种具有层次的数据结构
- 图，是一种更为复杂的数据结构
- 哈希函数与哈希表
- 位图，为n个二进制位的串，用于表示n项的状态

操作系统在计算机中的位置



OS与各种之间的关系

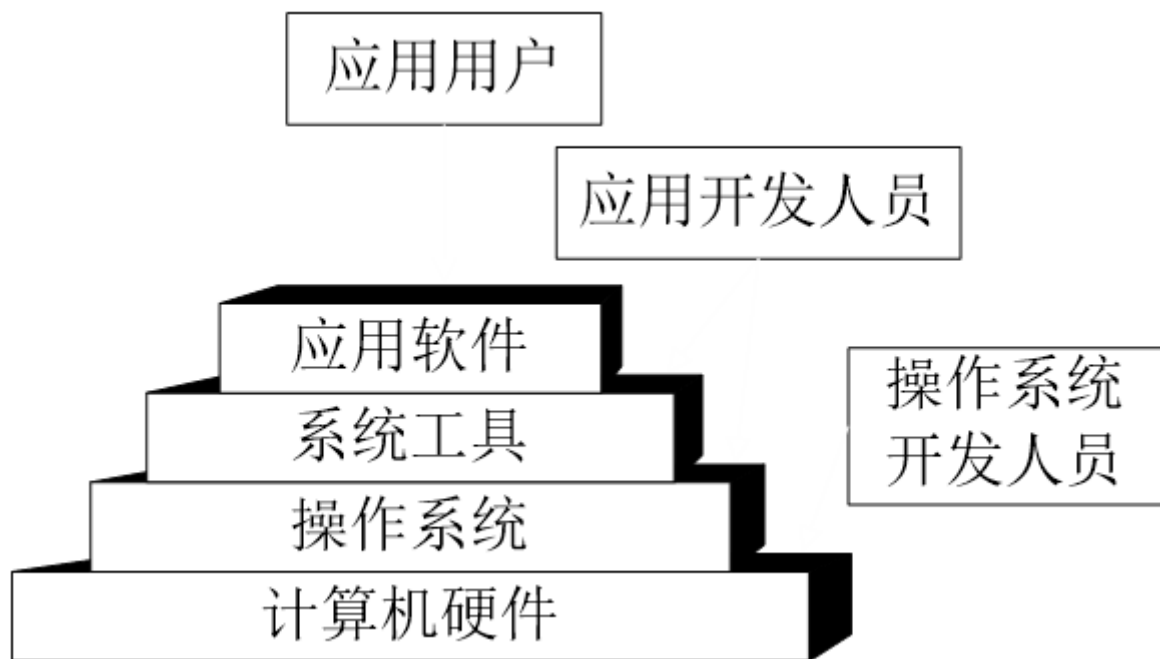
与硬件的关系：

- 控制CPU的工作
- 访问存储器
- 设备驱动、中断处理

与用户及应用程序的关系：

- 提供方便的用户界面
- 提供优质的服务

操作系统在计算机系统中的地位



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）

操作系统的目标和作用

目标

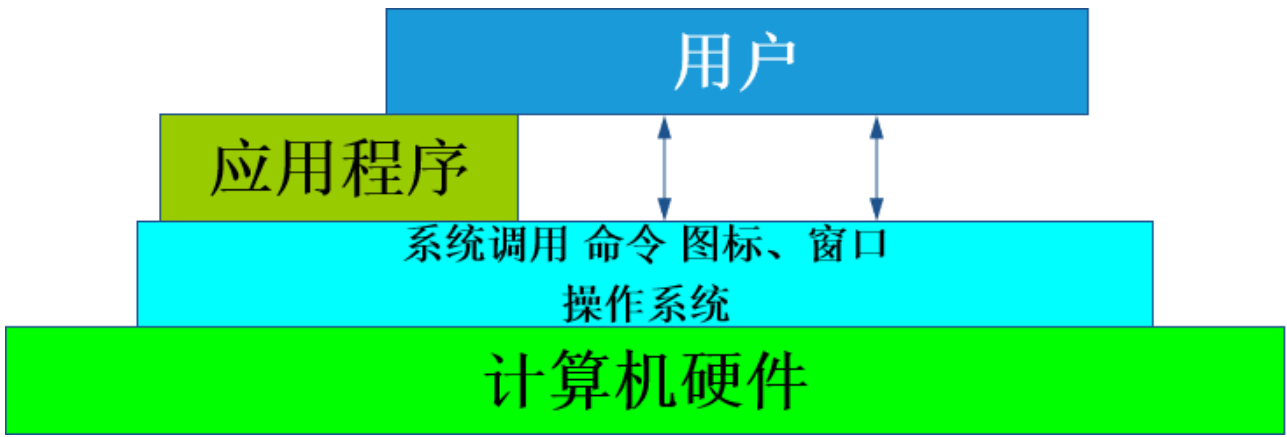
- 有效性：改善资源的利用率，提高系统吞吐量
- 方便性：使计算机系统使用起来更方便
- 可扩充性：能够不断适应发展的要求
- 开放性：使来自不同厂家的计算机和设备能够有效地协同工作，实现应用的可移植性和互操作性

作用

1. 操作系统作为用户与计算机硬件之间的接口

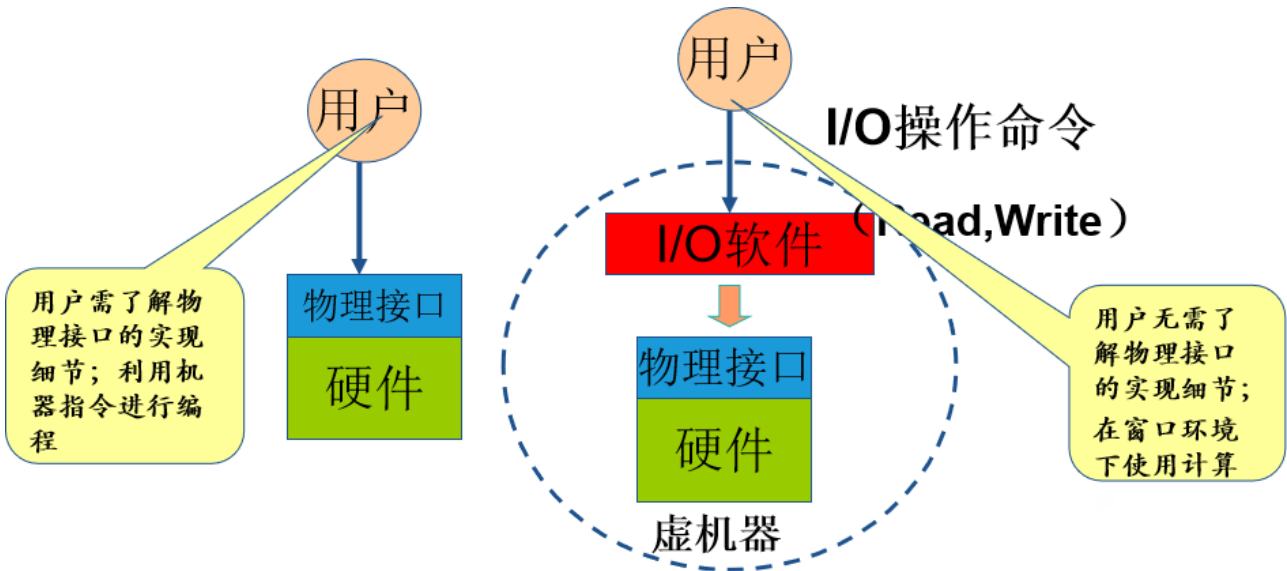
OS处于用户和计算机硬件系统之间，用户通过OS来使用计算机系统；

用户可以通过命令方式。系统调用方式使用计算机。



2.OS作为计算机系统的资源管理者

OS控制管理计算机资源的技术主要有3种，分别是资源复用，资源虚拟和资源抽象。当计算机上覆盖了OS后，便为用户提供了一台功能显著增强，使用更加方便，效率明显提高的虚拟计算机。



操作系统的定义

OS是控制和管理计算系统中各种硬件和软件资源，合理地组织

计算机工作流程，以及方便用户使用的一种软件。

OS的形成与发展

OS的形成

人工操作阶段

- 组成：主机运控部件、主存、输入设备、输出设备和控制台组成
- 特点：速度慢，规模小，外设少
- 缺点：用户独占全机；CPU等待人工操作

批处理操作阶段

为了减少人工操作所花的时间，提高资源利用率，人们首先为机器配备专门的操作员，程序员不再直接操作机器，减少操作失误。然后利用计算机系统软件代替操作员的部分工作，产生了最早的批处理操作系统。

基本思想

设计一个常驻主存的程序，称为监督程序。操作员有选择地把若干用户作业合成一批，安装在输入设备上，并启动监督程序。然后由监督程序自动控制这批作业运行。

执行过程

监督程序首先把第一道作业调入主存，并启动该作业，并将控制权交给作业，使之运行。第一道作业运行结束后再把控制器交给监督程序，监督程序再将下一道作业调入主存启动运行。如此下去，待一批作业全部处理结束后，操作员把作业运行结果一起交给用户。

优点

各作业之间的转换及运行完全由监督程序自动控制，从而减少部分人工干预，有效地缩短了作业运行前的准备

联机批处理

特点

在联机批处理系统中，慢速的输入输出设备和主机相连，作业的输入，调入内存以及结果输出都是在CPU的直接控制下进行的。

优点

解决了作业自动转换问题，从而减少了作业建立和人工操作的时间。

缺点

在作业的输入和执行结果的输出过程中，CPU仍处于停止等待状态，CPU的时间仍有很大的浪费，于是慢速的输入输出设备与快速的CPU之间形成了一对矛盾。

脱机批处理

特点

增加一台不与主机直接相连而专门用于与输入输出设备打交道的卫星机。卫星机又称为外围计算机，它不与主机直接相连，只与外部设备打交道。由于程序和数据的输入输出是在外围计算机的控制下完成的，或者说在脱离计算机的情况下进行的，故称为脱机输入输出。

优点

1. 主机和卫星机可以并行操作，二者分工明确，可以充分发挥主机的高速计算能力。
2. 利用外围机解决人工操作的缺点。
3. 提高I/O速度且减少了CPU的空闲时间。

执行系统阶段

通道的引入和中断计数的出现，这两项重大成果导致操作系统进入执行系统阶段

特点

借助于通道和中断技术，I/O的工作可在主机控制下完成。这时，原有的监督程序的功能扩大了，它不仅负责调度作业自动地运行。而且还要提供I/O控制功能。

这个发展了的监督程序常驻主存，称为执行系统。执行系统阶段是OS的初级阶段，它为操作系统的最终形成奠定了基础。

OS的完善

- 多道批处理系统
- 分时系统
- 实时系统
- 通用操作系统

OS的发展

- 单用户操作系统
- 网络操作系统
- 分布式操作系统
- 嵌入式操作系统

操作系统的分类

多道批处理系统

早期的批处理系统中只有一道作业在内存，系统资源的利用率仍然不高。为了提高资源利用率和系统吞吐量，引入多道程序设计技术，形成多道批处理系统

单道批处理系统

是最早出现的一种OS。严格来说，它只能算作OS的前身而并非现在人们所理解的OS

运行特征

- 单道性：内存中仅存放一个作业运行，即监督程序每次从磁带上只调入一道程序进入内存运行；
- 顺序性：磁带上的各道作业是顺序地进入内存，各道作业的完成顺序与他们进入内存的顺序，在正常情况下应完全相同，亦即先调入内存的先完成；

- 自动性：在顺利情况下，在磁带上的一批作业能自动地逐个地一次运行，而无需人工干预。

多道程序设计技术的含义

多道程序设计技术在计算机内存中同时放几道相互独立的程序，它们在管理程序的控制下相互穿插地运行。

多道程序设计的基本概念

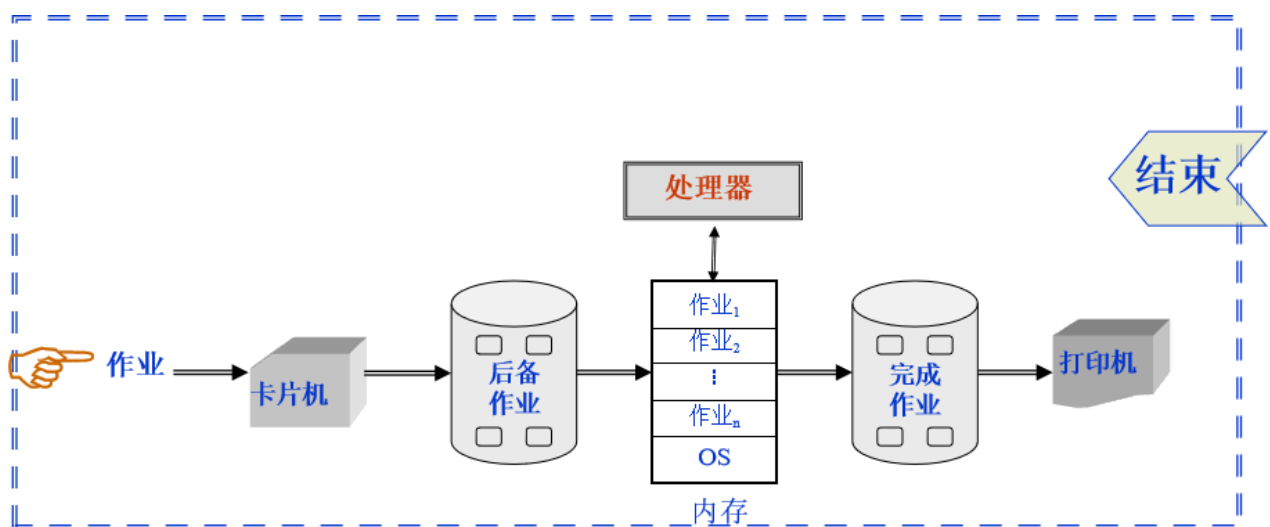
在该系统中，用户所提交的作业都先存放在外存上并排成一个队列，称为“后备队列”；然后，由作业调度程序按一定的算法从后备队列选择若干个作业调入内存，使它们共享CPU和系统中的各种资源。

在计算机内存中同时存放若干道已开始运行但尚未结束的程序，它们交替运行，共享系统中的各种硬、软件资源，从而使处理机得到充分利用。

运行特点

- 多道：内存中同时存放几个作业；
- 宏观上并行运行：都处于运行状态，但都未运行完；
- 微观上串行运行：各作业交替使用CPU；

工作原理



体现多道和成批特征，多道程序系统必备的硬件基础是中断技术。

优点

- 资源利用率高：CPU和内存利用率高；
- 作业吞吐量大：单位时间内完成的工作总量大。

缺点

- 用户交互性差：整个作业完成后或中间出错时，才与用户交互，不利于调试和修改。
- 作业平均周转时间长：短作业的周转时间显著增长，作业周转时间是指从作业进入系统开始，直到其完成并退出系统为止所经历的时间。

多道和单道批处理的比较

特征	单道	多道
内存使用	每一个作业	每次多个作业（充分利用内存）
作业次序	顺序，先进先出	无确定次序

多道程序系统和多处理系统的区别：前者是多个程序同时在内存中交替运行，后者指多个处理器。

道数

从表面上看，似乎道数越多越能提高效率，但是道数的多少绝不是任意的，它往往由系统的资源以及用户的要求而定。

计算利用率

1. 画甘特时序图，求出运行总时间和CPU所占用的时间。
2. 采用概率的方法计算。CPU的利用率= $1-p^n$, p 是一道程序等待I/O操作时间占其整个运行时间的比例，当主存中有 n 道程序时，则所有 n 道程序都是同时等待I/O的概率是 p^n , n 称多道程序的道数或度数。

多道批处理系统的特征

- 多道性：在内存中可驻留多道程序。
- 无序性：多个作业完成的先后顺序与他们进入内存的顺序之间并无严格的对应关系。
- 调度性：作业从提交给系统开始直至完成需要经历作业调度和进程调度。

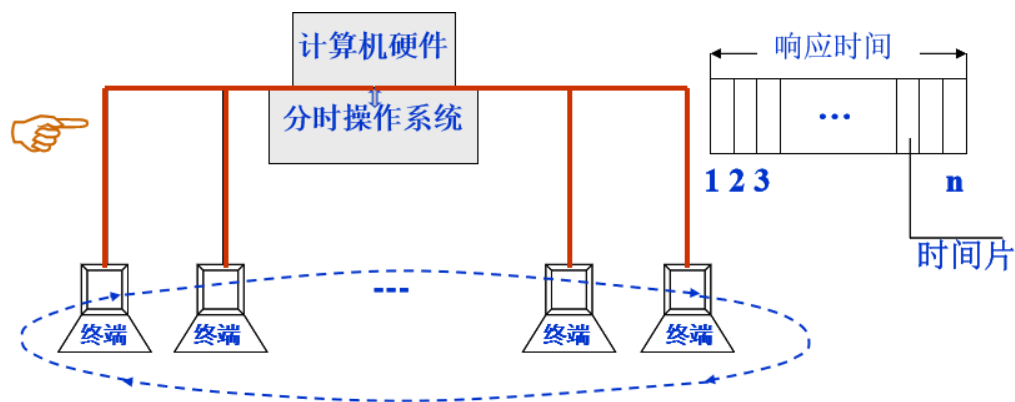
分时操作系统

批处理系统的目标是提高计算机系统的工作效率，但批处理系统的用户交互性差和作业平均周转时间长导致了分时系统的出现。

分时系统是指，在一台主机上连接了多个终端，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。

思想

- 采用时间片轮转的方法，同时为许多终端用户服务，对每个用户能保证足够快的响应时间，并提供交互会话的功能。
- 时间片：将CPU的时间划分成若干个片段，称为时间片，操作系统以时间片为单位，轮流为每个终端用户服务。
- 设计目标：对用户的请求及时响应，并在可能条件下尽量提高系统资源的利用率。



分时系统示意图

130

工作方式

- 一台主机连接了若干个终端
- 每个终端有一个用户在使用
- 交互式的向系统提出命令请求
- 系统接收每个用户的命令
- 采用时间片轮转方式处理服务请求
- 并通过交互方式在终端上向用户显示结果
- 用户根据上部结果发出下道命令

特点

- 多路性：同时有多个用户使用一台计算机。宏观上：是多人同时使用一个CPU，微观上：多个人在不用时刻轮流使用CPU
- 交互性：用户根据系统响应结果进一步提出新的请求（用户直接干预每一步）
- 独立性：每个用户各占一个终端，彼此独立操作，互不干扰。用户会感觉独占主机。
- 及时性：系统对用户提出的请求及时响应

优点

1. 人机交互性好
2. 共享主机
3. 用户独立性

实时操作系统

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该时间的处理，并控制所有实时任务协调一致地运行。

特点

及时性要求高，系统可靠性高。

分类

实时控制：通常是指以计算机为中心的生产过程控制系统，又称计算机控制系统。

实时信息处理：计算机及时接收从远程终端发来的服务请求，根据用户提出的问题对信息进行检索和处理，并在很短时间内对用户做出正确回答。

追求目标

- 对外部请求在严格时间范围内作出反应
- 高可靠性
- 安全性
- 完整性

实时系统与分时系统特征比较

特征	实时系统	分时系统
多路性	对多路现场进行采样与控制	为多个终端用户服务
独立性	独立采集信息与控制对象	各用户独立操作，互不干扰
实时性	以被控对象的可接受时间为准	以人可接受时间为准
交互性	仅与特定服务程序交互	与人类用户进项全方位交互
可靠性	高度可靠性	在人类可容忍的误差范围内可靠

微机操作系统的发展

单用户单任务操作系统

只允许一个用户上机，且只允许用户程序作为一个任务运行。如DOS

单用户多任务操作系统

只允许一个用户上机，但允许用户把程序分为若干个任务，使它们并发执行，从而有效地改善系统的性能。如Windows XP等。

多用户多任务操作系统

允许多个用户通过各自的终端使用同一台机器，共享主机系统中的各种资源，而每个用户程序又可进一步分为几个任务，使它们能并发执行，从而提高资源利用率和系统吞吐量。如Unix等。

网络操作系统

是在通常操作系统功能的基础上提供网络通信和网络服务功能的操作系统。

通常操作系统的功能是：

- 处理机管理
- 存储器管理
- 设备管理

- 文件管理

网络通信功能：通过网络协议进行高效、可靠的数据传输

网络资源管理：协调个用户使用

网络服务：文件和设备共享，信息发布

网络管理：安全管理、故障管理、性能管理等

互操作：直接控制对方，比交换数据更为困难

分布式OS

分户式计算机系统是以计算机网络为基础的计算机系统，包含多台处理机，每台处理完成系统中指定的一部分功能。从硬件上讲，它与计算机局域网没有任何区别，关键是软件。

分布式系统：处理和控制的分散（相对于集中式系统）

分布式系统是以计算机网络为基础的，他的基本特征是处理上的分布，即功能和任务的分布

分布式操作系统的所有系统任务可在系统中任何处理机上运行，自动实现全系统范围内的任务分配并自动调度各处理机的工作负载

分布式操作系统的特征

1. 是一个统一的操作系统

若干个计算机课相互协作共同完成一项任务

2. 资源进一步共享

3. 透明性

资源共享，分布对用户来讲是不知道的

4. 自治性

处于分布式系统的多个主机处于平等地位，无主从关系

5. 处理能力增强，速度更快，可靠性增强

嵌入式OS

是运行在嵌入式系统环境中，对整个嵌入式系统以及它所操作、控制的各种部件装置等等一切资源进行统一协调、调度、指挥和控制的系统软件。

操作系统的主要特性

并行性

两个或多个时间在同一时间间隔内发生。与并行不同，并行是指两个或多个事件在同一时刻发生。

在多道程序环境下，并行性是指在一段时间内宏观上有多个程序同时运行，微观上交替执行（在处理器情况下）。如果系统中有多个处理机，则这些可以并发执行的程序便可被分配到多个处理机上，实现并行执行，即利用每个处理机来处理一个可并发执行的程序，这样，多个程序便可同时执行。程序的静态实体是可执行文件，而动态实体是进程（或称作任务），并发指的是进程的并发。

进程：

在系统中你那个独立运行并作为资源分配的基本单位，它是由一组机器指令，数据和堆栈等组成的，是一个能独立运行的活动实体。

线程：

通常一个进程可以包含若干线程，他们利用进程所拥有的资源。在引入线程的OS中，通常都是把进程作为分配资源的基本单位。而线程作为独立运行和独立调度的基本单位。

共享性

多个线程共享有限的计算机系统资源。操作系统要对系统资源进行合理分配和使用。资源在一个时间段内交替被多个进程所用。

互斥共享方式（如打印机），资源分配后到释放前，不能被其他进程所用。

同时访问方式，（如可重入代码。磁盘文件）

虚拟性

虚拟性是指一个物理实体映射为若干个对应的逻辑实体（分时或分空间）。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。

时分复用技术

时分复用技术适用于数字信号的传输。由于信道的位传输速率超过每一路信号的数据传输速率，因此可将信道按时间分成若干片段轮换地给多个信号使用。每一时间片由复用的一个信号单独占用，在规定的时间内多个数字信号都可按要求传输到达，从而也实现了一条物理信道上传输多个数字信号。

空分复用技术

空分复用技术是指利用空间的分割实现复用的一种方式，将多根光纤组成成束是按空分复用，或者在同一根光纤中实现空分复用。

可以被虚拟的物理资源

计算机操作系统中可以被虚拟的物理资源包括虚拟处理机、虚拟I/O设备，虚拟磁盘和虚拟主存等

虚拟处理机

在虚拟处理机技术中，利用多道程序设计技术，为每道程序建立一个进程，让多道程序并发的执行，一次来分时使用一台处理机。此时，虽然系统中只有一台处理机，但它却能同时为多个用户服务，使每个终端用户都认为是有有一个处理机在为他们服务。利用多道程序设计技术，把一台物理上的处理机虚拟为多台逻辑上的处理机，在每台逻辑处理机上运行一道程序。我们把用户所感觉到的处理机称为虚拟处理机。

虚拟I/O

用在I/O设备的虚拟技术称为SPOOLing技术。通过虚拟设备技术可将一台物理I/O设备虚拟为多台逻辑上的I/O设备。并允许每个用户占用一台逻辑上的I/O设备，这样便可使原来仅允许在一段时间内由一个用户访问的设备（临界资源），变为在一段时间内允许多个用户同时访问的共享设备。例如，原来的打印机属于临界资源，而通过虚拟设备技术，可以把它变为多台逻辑上的打印机，供多个用户“同时”打印。

虚拟磁盘

通常在一台机器上只配置一台硬盘。我们可以通过虚拟磁盘技术将一台磁盘虚拟为多台虚拟磁盘，这样使用起来既方便又安全。虚拟磁盘技术也是采用例如空分复用方式，即它将硬盘划分为若干个卷，如1,2,3,4折4个卷，在通过安装程序将它们分别安装在C、D、E、F这个4个逻辑驱动器上。这样，机器上便有了4个虚拟磁盘。当用户要访问D盘中的内容时，系统便会访问卷2中的内容。另外，可以使用虚拟技术把磁盘虚拟为主存，分配给进程使用，作为主存的扩充。

虚拟主存

在单道程序环境下，处理机会有很多空闲时间，主存也会有很多空闲空间，显然，这会使处理机和主存的效率低下。如果说时分复用技术是利用处理机的空闲时间来运行其他程序，使处理机的利用率得以提高，那么空分复用则是利用存储器的空闲空间来存放其他的程序，以提高主存的利用率。但是，单纯的空分复用存储器只能提高主存的利用率，并不能实现在逻辑上扩大存储器容量的功能，必须引入虚拟存储技术才能达到此目的。而虚拟存储技术在本质上就是使主存时分复用。它可以使一道程序通过时分复用方式，在远小于它的主存空间中运行。

异步性

异步性也成不确定性，指进程的执行顺序和执行时间的不确定性。

多道程序设计环境下，程序按异步方式运行。

多个进程并发执行，“时走时停”，不可预知每个运行推进快慢，引发执行顺序与时间的不确定。

如果没有很好的同步机制，可能会导致程序执行结构不确定，不可再现。

相同输入与环境下多次运行结果下（当前操作系统中，都配置有很好的同步机制，保证程序多次运行的结构相同）

操作系统的功能

操作系统的主要任务

为多道程序的运行提供良好的环境，以保证多道程序有条不紊地、高效地运行，并能最大限度地提高系统中各种资源的利用率和方便用户的使用。

操作系统必须使用三种基本的资源管理技术才能达到目标，他们分别是资源复用或资源共享技术、虚拟技术和资源抽象技术。资源抽象技术用于处理系统的复杂性，解决资源的易用性。资源抽象软件对内封装实现细节，对外提供应用接口，使得用户不必了解更多的硬件知识，只通过软件接口即可使用和操作物理资源。操作系统中最基础和最重要的三种抽象是文件抽象。虚拟存储器抽象和进程抽象。

3中抽象之间存在一种包含关系，文件是对设备的抽象；虚拟存储器是对主存和设备的抽象；进程则是对处理机、主存和设备的首相和、进程是相对独立的自治单元，进程之间仅能通过内核所提供的的优先数目的原语或系统调用进行交互，操作系统在3种抽象的基础上能够很方便地控制程序的执行，调度并分配处理机资源。与进程抽象有关的所有工作称为进程管理。



处理机管理

处理管理的主要任务是：创建和撤销进程，对进程的运行进行协调，实现进程间的信息交换，以及按照一定的算法把处理机分配给进程。

功能

进程控制：创建、撤销、挂起、改变进程运行优先级等---主动改变进程的状态

进程同步：协调并发进程之间的推进步骤，以协调资源共享；共享有2种方式1.进程互斥2.进程同步

进程通信：进程之间传送数据，以协调进程间的协作； ---交换信息能力强，也可以用来协调进程之间的推进

作业和进程调度：作业和进程的运行切换，以充分利用处理机资源和提高系统性能。

存储器管理

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以及能从逻辑上扩充内存

功能

内存分配：为各个程序分配以及回收内存：

内存保护：保证进程间互不干扰、相互保密；如：访问合法性的检查。甚至要防止从“垃圾”中窃取其他进程的信息；

地址映射（变换）：进程逻辑地址到内存地址的映射。

内存扩充（覆盖、交换和虚拟存储）：提高内存利用率、扩大进程的内存空间；

设备管理

主要任务

完成用户提出的I/O请求，为用户进程分配所需的设备，提高CPU和I/O设备的利用率，提高I/O速度，方便用户使用I/O设备。

功能

缓冲管理：匹配CPU和外设的速度，提高两者的利用率（单缓冲区。双缓冲区和公用缓冲区）

设备分配：在多用户间共享I/O设备资源

设备处理：利用设备驱动程序完成对设备的操作

设备无关性：设备无关性也称设备独立性，是指应用程序独立于具体的物理设备。用户的程序不局限于某个具体的物理设备，这提高了用户程序的适应性，而且易于实现输入，输出的重定向。

文件管理

任务

对用户文件和系统文件进行管理以方便用户适应，并保证文件的安全性。

功能

文件存储空间的管理：解决五河存放信息，以提高空间利用率和读写性能。

文件和目录管理：文件的建立、读、写和目录管理等操作时文件系统管理的基本功能。文件系统管理负责根据各种文件的操作要求，完成所规定的任务。目录管理是为每个文件连理其目录项，并对众多的目录项加以有效组织，形成目录文件，以便实现文件按名存取。

文件共享和保护：文件共享是指多个用户可以使用同一个文件。为了防止用户多文件的非授权或越权访问，文件系统应该提供可靠保护措施，如采用口令、加密、存取权限等手段。

实现文件名到物理地址的映射：这种映射对用户是透明的，用户不必了解文件存放的物理位置和查找方法，只需指出文件名就可以找到相应的文件。这一映射是通过在文件说明部分中文件的物理地址来实现的。

提供方便的接口：为用户提供统一，方便的接口，主要是有关文件操作的系统调用，供用户编程时使用。

用户接口

目标

提供一个友好的用户访问操作系统的接口

分类

命令接口：为了便于用户直接或间接地控制自己的作业，操作系统向用户提供了命令接口。用户可通过该接口向作业发出命令以控制作业的运行。该接口又进一步分为联机用户接口、脱机用户接口和图形用户接口。

程序接口：供用户程序和系统程序调用操作系统功能。系统调用和高级语言库函数。

操作系统的运行环境

操作系统的运行机制

处理机的状态

在计算机系统中，根据执行程序的性质不同，CPU可以在两种不同的状态下工作，即核心态和用户态。核心态又称管态、系统态，是操作系统管理程序执行时机器所处的状态。它具有较高的特权，能执行包括特权指令在内的一切指令，能访问所有寄存器和存储区，且所占有的处理机是不允许被抢占的。用户态又称目态，是用户程序执行时奇迹所处的状态，是一种具有较低特权的执行状态，它智能执行规定的指令，访问指定的存储器和存储区，其所占有的处理机是可以被抢占的。

CPU的着两种状态是通过其内部的程序状态字寄存器表现出来的，在程序状态字寄存器中专门设置一个二进制位来表示CPU的状态，根据运行程序对资源和指令的使用权限不同而设置不同的CPU状态。

这两种状态可以相互转换，从用户态到核心态，可以通过系统调用、中断、异常或访管指令来实现；从核心态到用户态，可以通过中断返回指令和设置程序状态字的值来实现。

操作系统内核

操作系统中一些与硬件关联较紧密的模块（诸如时钟管理程序、中断处理程序、设备驱动程序处于操作系统最底层的程序）与运行频率较高的程序（诸如进程管理、存储管理和设备管理等）构成了操作系统的内核。

内核中指令工作在核心态。内核是计算机上配置的最底层的软件，是计算机功能的延伸。不同系统对内核的英译稍有区别，大多数操作系统内核包括4方面的内容

1. 时钟管理
2. 中断机制
3. 原语
4. 系统控制的数据结构及处理

中断与异常

中断也称为外中断，指来自CPU执行指令以外的事件发生，如设备发出的I/O结束中断，表示设备输入或输出处理已经完成，希望处理器能够想设备发出下一个输入或输出请求，同时让完成输入或输出后的程序继续进行。这一类重大关通常是与当前运行的程序无关。

中断可细分为硬中断和软中断。硬中断是硬件产生的，软中断是软件产生的。

异常也称为内中断、例外或陷入，指源自CPU执行指令内部的事件，如程序的非法操作码，地址越界，算数溢出，虚存系统的缺页及专门的陷入指令等引起的事件。对异常的处理一般要依赖于当前程序的运行现场，而且异常不能被屏蔽，一旦出现异常需要立即处理。

系统调用

所谓系统调用就是用户在程序中调用凑在哦系统所提供的一些子功能，系统调用可以被看做特殊的公共子程序。系统中各种共享资源都由操作系统统一管理，因此在用户程序中，凡是与资源有关的操作（如存储分配、进行I/O传输及管理文件等），都必须通过系统调用向操作系统提出服务请求，并由操作系统代为完成。通常，一个操作系统提供的系统调用命令有即使乃至上百条之多。这些系统调用按功能大致可分为设备管理、文件管理、进程控制、进程通信。内存管理、信息维护和安全管理等几类。

显然，系统调用运行在核心态。通过系统调用的方式来使用系统功能，可以保证系统的稳定性和安全性，防止用户随意更改或访问系统的数据或命令。系统调用是由操作系统提供的一个或多个子程序模块实现的。

在操作系统这一成眠上，我们关心的是系统核心态和用户态的群建实现和切换

用户态带核心态转换

- 用户程序要求操作系统提供服务，即系统调用
- 发生一次中断
- 用户程序中产生了一个错误状态
- 用户程序企图执行一条特权指令

从核心态转向用户态由一条指令实现，这条指令也是特权指令，一般是中断返回指令。

操作系统的性能指标

系统的RSA

RSA是指系统的可靠性（Reliability），可维修性（Serviceability）和可用性（Availability）三者的总称

系统吞吐量

吞吐量是指系统在单位时间内完成的作业数。在实际测量时应把各种类型的作业按一定的方式进行组合。

系统响应时间

响应时间是指从给定系统输入到系统响应的时间间隔。对于批处理系统，输入从用户提交作业时算起；对于分时系统，输入应从用户发出的中断命令时算起。

系统资源利用率

利用率是指在给定的时间内，系统内的某一资源，如CPU、内存和I/O设备等的实际使用时间所占的比例

可维护性

可维护性主要有两层含义，一是指在西戎运行过程中，不断排除系统设计中遗留下来的错误；二是对系统功能做某些修改或扩充，以适应新的环境或心得要求。

可移植性

可移植性的优劣可通过将操作系统移植到另一机型的工作量来衡量。工作量越少，系统性能越好。

方便用户

方便用户是指操作系统用户界面友好，使用灵活方便。

操作系统的体系结构

操作系统的体系结构就是操作系统的组成结构，操作系统的体系结构包括模块组合机构，层次结构和微内核结构。

模块组合结构

模块组合结构是早期操作系统及目前一些小型操作系统最常用的组织方式，它是基于结构化程序设计的一种软件结构设计方法。整个操作系统是一些过程模块的集合。系统中的每一个过程模块根据它们要完成的功能进行划分，然后按照一定的结构方式组合起来，协同完成整个系统的功能。在模块组合结构中，没有一致地系统调用界面，模块之间通过对外提供的接口传递信息，模块内部实现隐藏的程序单元，使其对其他过程模块来说是透明的。但是，随着功能的增加，模块间不加控制地相互调用和转移，以及信息传递方式的随意性，使系统存在一定的隐患。

层次结构

为了能让操作系统的结构更加清晰，使其具有较高的可靠性，较强的适应性，易于扩充和移植，在模块组合结构的基础上产生了层次结构的操作系统。所谓层次结构，即把操作系统划分为内核和若干模块（或进程），这些模块（或进程）按功能的调用次序排列成若干层次，各层之间只能是单向依赖或单向调用关系，即低层为高层服务，高层可以调用低层的功能，反之则不能，这样不但系统结构清晰，而且不构成循环调用。

采用层次结构的方法可以将操作系统各种功能分成不同的层析。最内层部分是计算机硬件本身提供的各种功能。同硬件紧挨着的是操作系统的内核，它是操作系统的核心层。内核包括中的中断处理、设备驱动、CPU调度及进程控制与通信等功能，其目的是提供一种进程可以存在和活动的的环境。内核以外依次是存储管理层、I/O管理层、文件管理层、作业管理层和命令管理层。它们提供各种资源管理功能并为用户提供各种服务。命令管理层是操作系统提供给用户的接口层，因而在操作系统的核心层。

微内核结构

微内核是以客户机/服务器体系结构为基础、采用面向对象技术的结构，能有效地支持多处理器，非常适用于分布式操作系统。微内核体系机构的基本思想是把操作系统中与硬件直接相关的部分抽取出来作为一个公共层，称为硬件抽象层。这个硬件抽象层的实质就是一种虚拟机，它向所有基于该层的其它层通过API（**Application Program Interface**，应用程序接口层）提供一系列标准服务。在微内核中只保留了处理机调度、存储管理和消息通信等少数几个组成部分，将传统系统内核中的一些组成部分放到内核之外来实现。