为什么要有操作系统？

用户几乎不可能使用裸机

计算机硬件只能识别0、1二值机器码

机器码直观性差。容易出错，难于交流

因此，通常在计算机硬件上会覆盖一层软件，以便用户使用计算机硬件

CPU是计算机硬件的核心，是计算机系统的心脏

操作系统则是计算机软件的核心，是计算机系统的大脑

操作系统是整个计算机的控制中心，是计算机中首要的、最重要的、最复杂的系统软件

什么是操作系统？

操作系统是管理计算机硬件的程序

在计算机用户和计算机硬件之间起媒介作用的一种程序

系统部件的抽象观点：硬件之上的第一层软件，是对硬件的首次扩充，又是其他软件运行的基础。

计算机系统部件：

硬件——提供基本的运算资源

操作系统——在各种应用程序和用户之间控制与协调对硬件的使用

应用程序——定义用户解决问题的资源使用方式（编译、数据库、视频程序、事务程序等）

用户——人、机、其他计算机

操作系统目标：

执行用户程序并使用户问题更易解决

使计算机系统更易使用。

以一种有效率的方式使用硬件

操作系统的其他定义：

资源分配者——管理和分配资源

控制程序——控制用户程序的运行和I/O设备的操作

内核——在全时运行的一个程序（其他的是应用），开机时最先打开的是操作系统，也是关机时最后关闭的

操作系统的作用：

计算机硬件、软件资源的管理者

用户使用计算机硬件、软件的接口

对用户——接待员、对系统——管家婆

操作系统（operating system）

操作系统发展：

os从无到有、从简单到复杂、完善

os随着计算机技术的发展而发展

为满足不同的需求，出现了多种类型的os

操作系统的发展过程：

发展动力——需求推动发展

1. 提高资源利用角度——为了提高计算机资源利用率和系统性能，从单道到多道、从集中到分布、从专用到泛在等
2. 用户角度——方便用户、人机交互
3. 技术角度——物理器件发展，CPU的位宽度（指令和数据）、快速补充、光器件等，以及计算机体系结构的不断发展：单处理机、多处理机、多核、计算机网络

Os的发展和计算机硬件技术、体系结构相关

第一代（1946年-1955年）：真空管时代，无操作系统

第二代（1955年-1965年）：晶体管时代、批处理系统

第三代（1965年-1985年）：集体电路时代，多道程序设计

第四代（1980年-至今）：大规模和超大规模集成电路时代，分时系统。现代计算机正在向着巨型、微型、并行、分布、网络化和智能化几个方面发展。

操作系统类型的发展：  
无操作系统——简单批处理系统——多道程序批处理系统——分时系统——实时系统——嵌入式系统——并行系统——分布式系统

1.手工操作 1946~50年代（真空管）——集中计算（计算中心）——计算机资源昂贵——用户独占全机

ENIAC计算机

运算速度：1000次每秒，数万个真空管，占地100平方米

没有程序设计语言（甚至没有汇编），更谈不上操作系统

工作方式：人工操作方式，用户是计算机专业人员；

工作方式： 编程语言：机器语言；

I/O纸带或卡片

缺点：

用户独占全机，独占操作系统的全部硬件资源，设备利用率很低

CPU等待用户：人工装入/卸取纸带或卡片

主要矛盾：

人机矛盾——人工操作方式与机器利用率的矛盾

CPU与I/O速度之间不匹配的矛盾

提高效率的途径：

批处理

脱机I/O

脱机I/O方式：

I/O工作在外围机/卫星机的控制下完成，或者说是在脱离主机的情况下进行。使用磁带作为输入输出的中介，这种具体的输入/输出不需要在主计算机上进行的方式称“脱机输入/输出”

2.单道批处理系统

（50年代末~60年代中）晶体管

把一批作业以脱机输入方式输入到磁带/磁鼓

利用磁带或磁鼓把任务分类编成作业顺序执行

每批作业由专门监督程序自动依次处理

运行特征：

顺序性磁带上的各道作业是顺序地进入内存，各作业的完成顺序与他们进入内存的顺序相同

单道性：

内存中仅有一道程序运行

自动性

优点：

减少了CPU的空闲时间，提高了主机CPU和I/O设备的利用效率，提高了吞吐量

缺点：

CPU和I/O设备使用忙闲不均

2.多道程序批处理系统

多道程序设计——60年代中~70年代中（集成电路）

多道——内存中同时存放几个作业，使之都处于执行的开始点和结束点之间的多个作业共享CPU、内存、外设等资源

目的——利用多道批出理提高资源的利用率

60年代通道和中断技术的出现

解决了输入输出等待计算的问题

通道——是一种专用部件，负责外部设备与内存之间信息的传输

中断——指主机或接到外界的信号（来自CPU外部或内部）时，立即中止原来的工作，转去处理这一外来事件，处理完后，主机又回到与原来工作点继续工作

在内存中同时有多个作业，CPU在其中切换

只要操作系统总是存在可执行的作业，CPU就永远不会因无事可干而闲着多道通过组织作业使得CPU总在执行其中一个作业，从而提高了CPU的利用率

运行特征：

多道性——内存中同时驻留多道程序并发执行，从而有效地提高了资源利用率和系统吞吐量

无序性——作业的完成顺序与它进入内存的顺序之间无严格的对应关系

调度性——作业调度、进程调度

优点——资源利用率高:CPU，内存，I/O设备

系统吞吐量大

缺点——无交互能力，用户响应时间长

作业平均周转时间长

多道程序对os特点的要求：  
存储管理——系统必须要为若干作业分配空间

CPU调度——系统必须要在就需作业中挑选

资源竞争和共享

设备分配

系统提供I/O程序

文件管理

1. 分时系统

在多道的基础上的新需求：  
多任务处理(多用户)——交互服务

70年代中期至今

分时是指多个用户分享同一台计算机，分时共享硬件和软件资源

实现方式：  
1.多用户分时——单个用户使用计算机的效率很低，因此允许多个应用程序同时在内存中，分别服务于不同的用户。有用户输入时由CPU执行，处理完一次输入后程序暂停，等待下一次用户输入—时走时停

2.前台和后台程序分时：后台程序不占用终端输入输出，不与用户交互—现在的图形用户界面（GUI），除当前交互的程序之外，其他程序均作为后台

通常按时间片分配：各个程序在CPU执行的轮换时间

主机——即使接受，即使处理

作业直接进入内存

每个作业一次只运行很短的时间

分时技术：把CPU的响应时间分成若干个大小相等（或不等）的时间单位，称为时间片（如100ms），每个终端用户获得CPU（获得一个时间片）后开始运行，当时间片到，该应用程序暂停运行，等待下一次运行

分时系统的特点

人机交互性好、共享主机：多个用户同时使用

分时系统的特点：

多路性——众多联机用户可以使用同一台计算机

独占性——各终端用户感觉到自己独占了计算机

交互性——用户与计算机之间可进行“会话”

及时性——用户的请求能在很短的时间内获得响应

在分时系统的基础上，操作系统的发展开始分化，如实时系统，通用（桌面）系统，网络系统、分布式系统等