

# 操作系统基本原理

## 1.操作系统概论

### 1. 操作系统的定义

#### 1. 计算机系统的组成

单向调用关系：外层软件必须使用实现定义好的方式调用内层软件或者硬件提供的服务(接口)

- 硬件：计算机硬件资源组成，CPU / 存储器 / 输入 / 输出
- 操作系统：计算机的管理和控制中心
- 实用程序：
  - 驻留在磁盘上，为用户提供良好的使用计算机的系统环境
  - 编译程序
  - 编辑程序
  - 调试程序
  - 连接装配和装入程序
- 应用程序：负责解决用户的实际问题

#### 2. 操作系统的定义和设计目标

- 计算机系统设计者：

操作系统是一系列程序模块组成的大的系统管理程序，使用各种管理策略，对计算机的软硬件资源进行管理和调度，提高资源的利用效率，硬件资源的管理控制程序
- 用户：

操作系统是用户和计算机系统之间的一个界面接口，使得系统变的容易维护和安全可靠，容错能力强，更加有效
- 操作系统的目标
  - 方便性：使用户方便简单的使用计算机系统
  - 有效性：是计算机系统高效可靠的运转
  - 便于操作系统的设计实现和维护

### 2. 操作系统的形成和发展

#### 1. 手工操作阶段(没有操作系统)

作业步：编写程序、编译程序、链接装配程序、装入可执行程序

1. 人工调度
2. 人工编排作业运行顺序

#### 2. 简单批处理

使用监控程序软件，监控程序常驻内存，操作员将多个作业按序成批地放在一个输入设备上，将CPU的控制权自动的从一个作业转换到另一个作业上去

##### 1. 联机批处理

主机的速度降低为慢速的输入输出设备

- 作业输入
- 计算
- 输出

##### 2. 脱机批处理

1. 为了提高CPU使用率，将CPU从慢速的外设中解放，引入了脱机批处理
2. 使用一台小型的卫星机专门用来控制外部设备的输入和输出
3. 本质：使用快速的输入输出设备代替慢速的设备，解放CPU

### 3. 多道批处理系统

根本目的：提高CPU的利用率，系统吞吐量达，利用率高

1. 以CPU为中心的体系结构转变成一主存为中心的体系结构

2. 新的设备和技术的出现

1. 通道

- 通道是一种慢于CPU但是价格便宜的硬件，比小型的卫星机更加的独立于CPU,控制输入输出的I/O处理机
- 通道连接着主存和外设，具有和主存直接交换数据的能力,使CPU和外设并行操作

2. 中断

当I/O设备完成传输后，通过中断机构向CPU报告完成情况

3. 缓冲

在主存设置缓冲区，来缓存用户的输入和输出，改善I/O设备和CPU之间速度不匹配的问题

3. 多道程序处理技术:

1. 主存中同时存放多个作业，使之同时处于运行状态共享系统资源的程序设计方法
2. 宏观上是并行的，微观上是并行的
3. 当其中一个任务正在中断(I/O)，可以转而处理其他的程序
4. 占用不同资源的程序可以并行执行，提高CPU利用率

4. 衡量批处理系统的性能指标

1. 资源利用率

在给定的时间内，系统的某一种资源的实际使用的时间所占的比例

2. 吞吐量

单位时间内系统所处理的信息量(单位时间内的处理的作业数)

3. 周转时间(平均周转时间)

作业从进入系统到作业退出系统所经历的时间，作用在系统的等待时间加运行时间

5. 多道技术的实现

1. 存储器的分配和保护

1. 保护操作系统
2. 独立隔离程序
3. 主存容量扩充技术

2. 处理机的管理和调度

3. 系统其他资源的管理和调度

### 4. 分时系统(多用户多任务操作系统)

■ 改进批处理系统

- 不能直接控制作业的运行：用户失去了对作业的控制权
- 作业周转时间过长
- 不能交互和快速响应

■ 分时

- 用户交互式的项系统发出请求，接受命令后使用时间片轮转的方式处理请求
- 并操作作为条件，并操作以分时共享系统资源作为基础
- 将CPU的单位时间划分成若干时间段，每个时间段又成为一个时间片，按照时间片轮转的方式CPU分配给每个用户使用
- 为了使被抢占的用户以后恢复运行，程序和数据将会被写入磁盘之后再装入新用户的程序和数据

■ 特点

- 同时性
- 独立性

- 交互性
- 及时性

响应时间：从用户发出请求到系统开始响应所需的时间

#### 5. 实时系统

1. 计算机可以对随机发生的外部事件及时的做出响应，强调反应速度
2. 首先看系统的相应的实时性，其次考虑资源的利用率
3. 特点
  1. 实时性：响应速度直接影响到现场控制的质量和和服务的质量
  2. 可靠性：硬件和软件必须有高可靠性，双工容错机制
  3. 确定性：
    1. 取决于系统响应中断的速度
    2. 系统是否存在足够的能力在要求的时间内完成指定的请求

#### 6. 嵌入式系统

##### 1. 定义

以应用为中心，以计算机技术为基础的软硬件可裁剪的，适应应用系统的功能，可靠性，等严格要求的专用的计算机系统，软件固化存储

##### 2. 功能

大多数的嵌入式系统都是一个多任务可抢占式的实时操作系统核心，只提供基本的功能

- 任务调度
- 任务同步和通信
- 主存管理
- 时钟管理

#### 3. 操作系统的功能，服务，特性

##### 1. 操作系统的功能

1. 处理机管理：处理机的作业的调度的策略，提高系统的资源的利用率(进程表)
2. 存储器管理：对计算机主存的管理，主存是处理CPU之后的有一个宝贵的资源(存储表)
3. 设备管理：输入 / 输出设备的管理和控制(设备表)
4. 文件管理：信息管理和文件系统(文件表)

##### 2. 服务

1. 用户接口
2. 执行程序
3. I/O操作
4. 文件系统操作
5. 通信服务
6. 错误检测和处理
7. 资源分配
8. 保护系统

##### 3. 特性

1. 并发性：增强计算机系统的处理能力采用的时间上重叠的技术
2. 共享性：是并发的基础，没有共享无法并发
3. 虚拟性：为了共享资源,将硬件物理实体虚拟成对应的逻辑上的多个实体
4. 异步性：系统允准的未知性，会给系统带来危险，操作系统必须要有良好的容错性来避免异步的危险

#### 4. 操作系统的进一步发展

##### 1. 个人计算机系统

##### 2. 多处理机操作系统

1. 非对称个多处理机操作系统：

- 一个处理机运行操作系统其他的处理机运行用户程序，前者为后者实现处理及分配和任务的调度
- 主从模式

## 2. 对称多处理机操作系统

- 操作通可以安排在任何一个处理机上运行，或者同时在多个处理机上运行
- 处理机共享主存和I/O设备
- 相对有更多的优势
  - 增加了系统的吞吐量：多个作业可以在任何一个处理机上运行
  - 可靠性：一个损坏只是性能降低

## 3. 网络操作系统

- 网络操作系统是为了计算机网络配置的操作系统，网络中的多台主机都有独立的操作系统，网络操作系统将他们有机的联系起来
- 为网络中的其他中级提供通信和网络资源的共享
- 功能
  - 处理机管理
  - 存储器管理
  - 设备管理
  - 文件管理
  - 网络管理
- 结构模式
  - C / S：服务器是瓶颈
  - 对等模式：对等无控制中心

## 4. 分布式操作系统

- 将多个分散的计算机经过互联网连接成为同意的计算机系统
- 每个计算机系统高度自制和相互配合
- 要求联网的多机存在统一的操作系统
- 例子：机群

## 5. 用户和操作系统的接口

### o 接口

- 操作接口
  - 命令行
  - GUI
- 编程接口：系统调用

用户程序和操作系统内核的接口，经常以API形式出现

将从用户态切换到核心态执行相应的系统调用，然后返回用户程序继续执行

本质：软中断，通过中断向量传递操作系统的服务程序，切换到核心态(陷入)

- 进程控制
- 文件管理
- 设备管理
- 其他服务

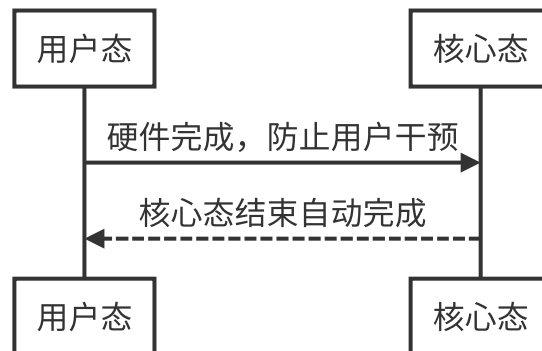
### o CPU的运行模式

- 为了防止用户程序破坏操作操作系统或者直接处理硬件，必须严格区分操作系统代码和用户代码
- 处理机状态字(PSW)

添加执行方式位

- 00：核心态,管理态，权限态
  - 可以使用特权指令，不允许用户使用

- 操作系统大部分运行在核心态，可以执行处理机的全部指令集，访问所有的寄存器和存储区
- 核心态为用户态服务
- 11：用户态
  - 用户程序
  - 只允许使用非特权指令集，访问特定的寄存器和存储区，否则会被操作系统故障终止



## 6. 操作系统的运行环境

### 1. 概念

#### 1. 中断：

1. 中断出现，CPU响应中断，中断处理程序调用加入当前的进程的用户核心栈
2. 外部设备向CPU发出的硬件信号，异步发生，可以屏蔽
3. 优先级
  1. 处理机优先级：指处理机当前正运行程序的中断响应级别
  2. 不同中断有不同优先级。处理高级中断时会临时屏蔽低级中断

#### 2. 异常：

1. 程序产生的同步发生的不可屏蔽的
2. 发生时机
  1. 程序错误
  2. 系统调用

## 7. 操作系统的设计规范

### 1. 系统效率

1. 资源利用率
2. 吞吐量
3. 周转时间 / 响应时间

### 2. 系统可靠性

### 3. 可移植性

### 4. 可伸缩性：对软硬件资源的适应能力

### 5. 兼容性

### 6. 安全性

## 2. 纲要总结

### 1. 早期操作系统的设计目标？

#### 操作系统的目标

- 方便性：使用户方便简单的使用计算机系统

- 有效性：是计算机系统高效可靠的运转
  - 便于操作系统的设计实现和维护
2. 操作系统是资源管理程序，它管理系统中的什么资源？      ???????
- 1. 处理机管理：处理机的作业的调度的策略，提高系统的资源的利用率(进程表)
  - 2. 存储器管理：对计算机主存的管理，主存是处理CPU之后的有一个宝贵的资源(存储表)
  - 3. 设备管理：输入 / 输出设备的管理和控制(设备表)
  - 4. 文件管理：信息管理和文件系统(文件表)
3. 为什么要引入多道程序系统?它有什么特点？
- 1. 引入多道程序系统  
提高CPU的利用率，系统吞吐量达，利用率高
  - 2. 特点
    - 1. 主存中同时存放多个作业，使之同时处于运行状态共享系统资源的程序设计方法
    - 2. 宏观上是并行的，微观上是并行的
    - 3. 当其中一个任务正在中断(I/O)，可以转而处理其他的程序
    - 4. 占用不同资源的程序可以并行执行，提高CPU利用率
4. 叙述操作系统的基本功能？
- 1. 处理机管理：处理机的作业的调度的策略，提高系统的资源的利用率(进程表)
  - 2. 存储器管理：对计算机主存的管理，主存是处理CPU之后的有一个宝贵的资源(存储表)
  - 3. 设备管理：输入 / 输出设备的管理和控制(设备表)
  - 4. 文件管理：信息管理和文件系统(文件表)
5. 批处理系统、分时系统和实时系统各有什么特点?各适合应用于哪些方面？
- 1. 批处理
    - 1. 优点: 系统吞吐量大，资源利用率高。缺点: 用户与作业无法交互，作业平均周转时间较长。
    - 2. 适合计算量大、自动化程度高的成熟作业
  - 2. 实时系统
    - 1. 实时，可靠，确定
    - 2. 实时过程控制，实时信息处理
  - 3. 分时系统
    - 1. 同时，独立，交互，及时
    - 2. 多用户系统，大型数据库查询
6. 操作系统的特性？
- 并发，共享，虚拟，异步
7. 衡量OS的性能指标有哪些？什么是吞吐量、响应时间和周转时间？
- 1. 资源利用率,吞吐量,周转时间 / 响应时间
  - 2. 概念
    - 1. 吞吐量：单位时间处理的作业个数(信息量)
    - 2. 周转时间：作业进入系统到作业退出系统的所有功能时间
    - 3. 响应时间：从用户发出请求到系统开始响应所需的时间
8. 什么是嵌入式系统？
- 以应用为中心，以计算机技术为基础的软硬件可裁剪的，适应应用系统的功能，可靠性，等严格要求的专用的计算机系统，软件固化存储
9. 什么是对称多处理？它有什么好处？
- 操作通可以安排在任何一个处理机上运行，或者同时在多个处理机上运行
  - 处理机共享主存和I/O设备
  - 相对有更多的优势
    - 增加了系统的吞吐量：多个作业可以在任何一个处理机上运行
    - 可靠性：一个损坏只是性能降低

10. 为了实现系统保护，CPU通常有哪两种工作状态？各种状态下分别执行什么程序？什么时候发生状态转换？状态转换由谁实现的？
1. 核心态，用户态
  2. 操作系统的核心程序，用户程序和部分的操作系统的程序
  3. 中断或者异常
  4. 用户态到核心态硬件，核心态到用户态自动
11. 什么是系统调用？什么是特权指令？特权指令执行时，CPU处于哪种工作状态？
1. 用户程序和操作系统内核的接口，经常以API形式出现
  2. 关系系统全局的指令,很重要的指令，只允许操作系统使用
  3. 核心态
12. 操作系统通常向用户提供哪几种类型的接口?其主要作用是什么？
1. 操作接口和编程接口
  2. 接口作用
    1. 操作接口：可以利用接口来组织和控制作业的运行
    2. 系统接口：编程人员利用，向操作系统提出资源请求和功能服务