# 操作系统简答题（知乎）

一、计算机系统概述

1、操作系统的功能是什么？

处理机管理、内存管理、文件管理和设备（I/O）管理。

2、试解释并发和并行的区别。

并发是指多个事件在同一时间间隔内发生，并行是指多个事件在同一时刻发生。

例如你在9: 00到9: 30仅吃面包，9: 30到10: 00仅写字，那么在9: 00—10: 00这一时间段内，吃面包和写字就是并发的；如果你在9: 00—10: 00左手拿着面包吃，右手写字，那这两个动作就是并行的。

3、什么是管态和目态？如何从目态切换到管态？

管态又叫系统态、核心态。CPU在管态下有权限执行计算机的任何指令，其资源访问不受限制。操作系统内核程序处于管态。

目态又叫用户态。CPU处于目态时，程序只能执行非特权指令，不能直接使用系统资源，并且只能访问用户程序自己的存储空间。用户编写的程序处于目态。

区分管态和目态是为了保护系统程序。

通过系统调用（使用访管指令），可以从目态切换到管态。此外，程序产生异常或中断时（如除以0，缺页，I/O中断），也会切换到内核态。

二、进程管理

1、进程的基本状态切换有哪些？

就绪态→运行态。进程被调度后，分配CPU时间片，并使其运行运行态→就绪态。进程时间片用完后，让出CPU，转为就绪态运行态→阻塞态。进程请求某一资源或等待某一事件发生（如I/O完成），使自身阻塞阻塞态→就绪态。进程等待的事件到来时，如I/O操作结束，转为就绪态

2、进程和线程的区别是什么？

进程是系统资源分配的基本单位，线程是CPU调度的基本单位。

线程可看作一种“轻量级线程”。线程自己不拥有系统资源，只拥有一点儿在运行中必不可少的资源，它与所属进程的其他线程共享进程的系统资源。

引入线程的目的是为了减小程序在并发执行时的时空开销，提高操作系统的并发性能。

3、进程之间的通信方式有哪些？

1）共享存储

低级方式：P/V操作，基于数据结构高级方式：开辟共享内存，基于存储区

2）消息传递

直接通信：把消息挂到接收进程的消息队列信箱通信：通过中介者转发，类似电子邮件

3）管道通信：使用特殊的pipe文件，一方写入，一方读取

4、常用的调度方式和算法有哪些？

分为高级调度（作业调度）、中级调度（内存调度）、低级调度（进程调度）。

先来先服务（FCFS）、短作业优先（SJF）、时间片轮转优先级调度：分为抢占式和非抢占式高响应比优先：响应比=(等待时间+要求服务时间)/要求服务时间多级反馈队列调度：一个时间片结束后，未完成的进程转入低一级的就绪队列

5、死锁的产生条件是什么？如何解决死锁？

死锁产生的必要条件：

互斥：资源在同一时间只能被一个进程占有非剥夺：进程在使用完资源之前，不可被抢占请求与保持：进程保持已有的资源，又去申请新的资源循环等待：存在循环等待链，使得每个进程的资源都被下一进程所请求

解决死锁的方法：

预防死锁：破坏死锁产生的必要条件，使其无法发生避免死锁：动态分配资源时，用一些算法（如银行家算法）防止系统进入不安全状态死锁的检测与解除：检测到死锁发生后，采取措施（剥夺资源、终止进程）解除之

三、内存管理

1、为什么要引入动态重定位？如何实现？

在程序执行过程中，需要将要访问的数据的逻辑地址转换成物理地址。

具体实现方法是增加一个重定位寄存器，装入程序在内存中的起始地址，真正的地址为相对地址与重定位寄存器中的起始地址相加之和，从而实现动态重定位。

2、内存分区时有哪些策略？有何优缺点?

首次适应算法：按地址从大到小排序，分配第一个符合条件的分区循环首次适应算法：在首次适应的基础上，每次从上一次结束的位置开始查找最佳适应算法：按空间从小到大排序，分配第一个符合条件的分区最坏适应算法：按空间从大到小排序，分配第一个符合条件的分区

优缺点：理解原理后，从碎片的产生、是否缺乏大空闲区、排序性能等考虑

3、页面置换算法有哪些？什么是抖动？

先进先出置换算法（FIFO）：需要替换页面时，将最早调入的页面调出最近最久未使用置换算法（LRU）：记录每个页面上次被访问以来经历的时间，需要替换页面时，将最长时间未被访问的页面调出最近未用算法（CLOCK、NRU）：可看作LRU的简化版。为每个页面设置一个使用位，需要替换页面时，将使用位为1的置为0，将第一个使用位为0的页面调出

抖动：是指在页面替换过程中，刚刚调入的页面又被调出，刚刚调出的页面很快又被调入，发生频繁的页面调度行为。

4、其他概念（比较简单，略）

分页和分段、局部性原理、虚拟存储器

四、文件管理

1、常用的磁盘调度算法有哪些？

先来先服务（FCFS）最短寻找时间算法（SSTF）：总是处理离当前磁头最近的磁道请求，会产生“饥饿”现象扫描算法（电梯算法，SCAN）：选择在磁头当前移动方向上最近的磁道请求进行处理循环扫描算法（C-SCAN）：在SCAN的基础上，规定磁头单向移动。磁头到达最远端的请求后，快速返回起始端

五、输入输出管理

1、I/O控制方式有哪些？

程序查询：CPU循环检查外设状态，与外设串行工作，利用率极低中断驱动：I/O设备通过中断请求CPU服务，例如键盘设备DMA（直接存储器访问）：在I/O设备与内存之间开辟直接的通路，以数据块为基本单位进行传送，数据传送过程由DMA控制器完成，仅在传送开始和结束需要CPU干预I/O通道：DMA方式的发展。配备专用的I/O通道处理机，需要的CPU干预更少

2、什么是SPOOLing技术？举一个实例。

SPOOLing技术又称“假脱机技术”，是在通道技术和多道程序设计基础上产生的，它是一种将独占设备改造成共享设备的技术。

将一台独享的打印机改造成可供多个用户共享的打印机，是SPOOLing的典型应用。

具体做法是：系统对于用户的打印输出，先在输出井申请一个空闲盘块区，并将要打印的数据送入其中，然后将此作业挂在打印队列上。若打印机空闲，输出程序从打印队列中取出作业，将要打印的数据从输出井传送到内存缓冲区，再进行打印。