# 操作系统总复习及相关习题

## 第一章   引论

### 名词解释

**1操作系统**

操作系统是管理和控制计算机系统内各种硬件和软件资源，有效地组织多道程序运行的系统软件（或程序集合），是用户与计算机之间的接口。

**2管态**

当执行操作系统程序时，处理机所处的状态

**3目态**

当执行普通用户程序时，处理机所处的状态。

**4多道程序设计**

在这种设计技术下，内存中能同时存放多道程序，在管理程序的控制下交替的执行。这些作业共享CPU和系统中的其他资源。

**5并发**

是指两个或多个活动在同一给定的时间间隔中进行。它是宏观上的概念。

**6并行**

是指两个或多个活动在同一时刻同时执行的情况。

**7吞吐量**

在一段给定的时间内，计算机所能完成的总工作量。

**8分时**

就是对时间的共享。在分时系统中，分时主要是指若干并发程序对CPU时间的共享。

**9实时**

表示“及时”或“既时”。

**10系统调用**

是用户在程序中能以“函数调用”形式调用的、由操作系统提供的子功能的集合。每一个子功能称作一条系统调用命令。它是操作系统对外的接口，是用户级程序取得操作系统服务的唯一途径。

**11特权指令**

指指令系统中这样一些指令，如启动设备指令、设置时钟指令、中断屏蔽指令和清内存指令，这些指令只能由操作系统使用。

**12命令解释程序**

其主要功能是接收用户输入的命令，然后予以解释并且执行。

**13脱机I/O**

是指输入/输出工作不受主机直接控制，而由卫星机专门负责完成I/O，主机专门完成快速计算任务，从而二者可以并行操作。

**14联机I/O**

是指作业的输入、调入内存及结果输出都在cpu直接控制下进行。

**15资源共享**

是指计算机系统中的资源被多个进程所功用。例如，多个进程同时占用内存，从而对内存共享；它们并发执行时对cpu进行共享；各个进程在执行过程中提出对文件的读写请求，从而对磁盘进行共享等等。

**简答题**

**1什么是操作系统？它的主要功能是什么？**

答：操作系统是控制和管理计算机系统内各种硬件和软件资源，有效地组织多道程序运行的系统软件（或程序集合），是用户与计算机之间的接口。操作系统的主要功能有5个方面，即存储管理、处理机管理、设备管理、文件管理和用户接口。

**2推动操作系统形成和发展的主要动力是什么？**

答：推动操作系统发展的因素很多，主要可归结为两大方面：硬件技术更新和应用需求扩大伴随计算机器件的更新换代和计算机体系结构的发展，促使操作系统的性能和结构有了显著发展。 应用需求促进了计算机技术的发展，也促进了操作系统的不断更新升级。

**3操作系统的基本特征是什么？**

答：操作系统的基本特征是并发、共享和不确定。并发性是指两个或多个活动在同一给定的时间间隔中进行；共享是指计算机系统中的资源被多个进程所共用；不确定性是指系统中各种事件发生顺序的不可预测性。

**4多道程序和多重处理有何区别？**

答：多道程序是作业之间自动调度执行、共享系统资源，并不是真正的同时执行多个作业；而多重处理系统配置多个cpu，能真正同时执行多道程序。要有效使用多重处理，必须采用多道程序设计技术，而多道程序设计原则上不一定要求多重处理系统的支持。

**5试说明多道程序设计和多任务系统之间的关系**

答：多道程序设计是利用外设与cpu能够并行处理的特性，在主存同时存放多个程序，使之在系统中交叉地使用cpu，从而提高系统资源的利用率。而多任务系统主要指多进程交叉使用cpu。多道程序隐含了多任务处理，但多任务系统中不一定有多道程序。因为一个程序也可以采用多任务处理机制。

**6不同类型的操作系统提供不同的功能。假定有如下的应用环境，请你为它们选择适合的操作系统。**

**（1）飞机的导航，（2）办公自动化系统，（3）航空订票系统，（4）复杂的科学计算，（5）图书检索系统**

答：（1）飞机的导航系统，应采用硬实时操作系统

（2）办公自动化系统，应采用分时操作系统

（3）航空订票系统，应采用软实时操作系统

（4）复杂的科学计算，应采用批处理系统

（5）图书检索系统，应采用软实时操作系统

**7什么是批处理系统，它有什么特征？**

答：批处理系统：操作员把用户提交的作业分类，把一批作业编成一个作业执行序列，由专门编制的监督程序自动依次处理。其主要特征是：用户脱机使用计算机、成批处理、多道程序运行。

**8什么是分时系统，它有什么特征？**

答：分时系统：把处理机的运行时间分成很短的时间片，按时间片轮转的方式，把处理机分配给各进程使用。其主要特征是：交互性、多用户同时性、独立性。

**9什么是实时系统？它有什么特征？**

答：实时系统：在被控对象允许时间范围内做出响应 。其主要特征是：对实时信息分析处理速度要比进入系统快、要求安全可靠、资源利用率低。

**10什么是处理机的核心态和用户态？为什么要设置这两种不同的状态？**

答：当执行操作系统程序时，处理机处于核心态。它有较高的特权，可以执行所有的指令，包括一般用户程序中不能使用的特权指令，从而能对所有寄存器和内存进行访问，启动i/o操作等。

用户程序是在用户态下执行，它的权限较低，只能执行指令集中非特权指令。（2分）

设置这两种不同状态的目的是为了保护操作系统程序（特别是其内核部分），防止受到用户程序的损害。

**11系统调用与过程调用在功能及实现上有什么相同点和不同点？**

答：相同点：两者都由程序代码构成，可直接用高级程序设计语言（如C,C++和Perl语言）来编制；使用方式相同——以函数调用的形式出现，调用时传送参数。

    不同点：①代码层次不同，过程调用不属于操作系统的一部分，而系统调用是操作系统的一部分。②运行状态不同。过程调用只能在用户态下运行，不能进入核心态，而系统调用是在核心态下运行的。③进入方式不同。过程调用在用户程序中调用，并直接在用户空间内执行；而系统调用可以在用户程序中调用，但是在用户程序中执行到系统调用时，会产生异常事件。实现处理机状态从用户态到核心态的转变，从而进入操作系统核心空间去执行系统调用的代码。

**12试说明特权指令和系统调用之间的区别与联系。**

答：特权指令是一类只能在核心态下执行的机器指令。而系统调用不是机器指令，它往往以函数调用的形式出现，实现操作系统提供的子功能，它是操作系统与用户的编程接口 。在用户程序中可以使用系统调用来获得操作系统服务，在系统调用代码中可以使用特权指令

## 第二章  进程和线程

### 名词解释

**1顺序性**

是指顺序程序所规定的每个动作都在上个动作结束后才开始的特性。

**2封闭性**

是指只有程序本身的动作才能改变程序的运行环境。

**3可再现性**

是指程序的执行结果与程序运行的速度无关。

**4进程**

程序在并发环境中的执行过程。

**5互斥**

在逻辑上本来完全独立的进程，由于竞争同一个资源而产生的相互制约的关系。

**6同步**

是指进程间共同完成一项任务时直接发生相互作用的关系。也就是说，这些具有伙伴关系的进程在执行次序上必须遵循确定的规律。

**7临界资源**

一次仅允许一个进程使用的资源。

**8临界区**

在每个进程中访问临界资源的那段程序。

**9线程**

线程是进程中实施调度和分派的基本单位。

**10管程**

管程是一种高级同步机制，一个管程定义一个数据结构和能为并发进程在其上执行的一组操作，这组操作能使进程同步和改变管程中的数据。

**11进程控制块**

进程控制块是进程存在的唯一标识，它保存了系统管理和控制进程所必须的信息，是进程动态特性的集中表现。

**12原语**

指操作系统中实现一些具有特定功能的程序段，这些程序段的执行过程是不可分割的，即其执行过程不允许被中断。

**13就绪态**

进程已经获得了除cpu之外的全部资源，等待系统分配cpu，一旦获得cpu，进程就可以变为运行态。

**14运行态**

正在cpu上执行的进程所处的状态。在单cpu系统中，任何时候最多只能有一个进程处于运行状态。

**15阻塞态**

又称等待态，指正在运行的进程因等待某个条件发生而不能运行时所处的状态。处于阻塞态的进程在逻辑上是不能运行的，即使cpu空闲，它也不能占用cpu。

**16进程通信**

是指进程间的信息交换。

**17同步机制**

同步机构是负责处理进程之间制约关系的机制，即操作系统中负责解决进程之间协调工作的同步关系（直接制约关系），以及共享临界资源的互斥关系（间接制约关系）的执行机构。

**简答题**

**1在操作系统中为什么要引入进程概念？**

答:  由于多道程序并发执行时共享系统资源，共同决定这些资源的状态，因此系统中各程序在执行过程中就出现了相互制约的新关系，程序的执行出现“走走停停”的新状态。用程序这个静态的概念已不能如实反映程序并发执行过程中的这些特征。为此，人们引入了“进程（Process）”这一概念来描述程序动态执行过程的性质。

进程和程序是两个完全不同的概念。然而，进程与程序之间存在密切关系，进程的功能是通过程序的运行得以实现的，进程活动的主体是程序。进程不能脱离开具体程序而独立存在。

**2有人说，一个进程是由伪处理机执行的一个程序，这话对吗？为什么？**

答：对。

因为伪处理机的概念只有在执行时才存在，它表示多个进程在单处理机上并发执行的一个调度单位。因此，尽管进程是动态概念，是程序的执行过程，但是，在多个进程并行执行时，仍然只有一个进程占据处理机执行，而其他并发进程则处于就绪或等待状态。这些并发进程就相当于由伪处理机执行的程序。

**3试比较进程和程序的区别**

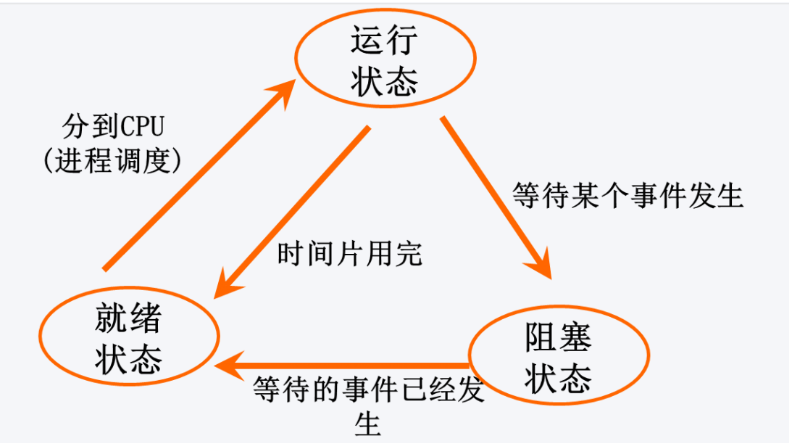
答：（1）进程是一个动态的概念，而程序是一个静态的概念，程序是指令的有序集合，无执行含义，进程则强调执行的过程。

（2）进程具有并行特征（独立性、异步性），程序则没有。

（3）不同的进程可以包含同一个程序，同一程序在执行中也可以产生多个进程。

**4进程的基本状态有哪些？试描绘进程状态转换图。**

答：进程至少有三种基本状态：运行状态、就绪状态和阻塞状态（或等待状态） 。进程状态转换如下图：



**5并发进程间的制约有哪两种？引起制约的原因是什么？**

答：并发进程所受的制约有两种：直接制约和间接制约。

直接制约是由并发进程相互共享对方的私有资源所引起的；间接制约是由竞争共有资源而引起的。

**6什么是进程间的互斥？什么是进程间同步？**

答：进程间的互斥是指：一组并发进程中的一个或多个程序段，因共享某一共有资源而导致它们必须以一个不许交叉执行的单位执行，即不允许两个以上的共享该资源的并发进程同时进入临界区。

进程间的同步是指：异步环境下的一组并发进程因直接制约相互发送消息而进行相互合作、相互等待，是各进程按一定的速度执行的过程。

**7什么是临界区和临界资源？进程进入临界区的调度原则是什么？**

答：临界资源——一次仅允许一个进程使用的资源

临界区——在每个进程中访问临界资源的那段程序

一个进程进入临界区的调度原则是：

① 如果有若干进程要求进入空闲的临界区，一次仅允许一个进程进入

② 任何时候，处于临界区内的进程不可多于一个。如已有进程进入自己的临界区，则其他所有试图进入临界区的进程必须等待

③ 进入临界区的进程要在有限的时间内退出，以便让其他进程能及时进入自己的临界区

④ 如果进程不能进入自己的临界区，则应让出cpu，避免进程出现“忙等”现象.

**8简述信号量的定义和作用。P，V操作原语是如何定义的？**

答：信号量一般是由两个成员组成的数据结构，其中一个成员是整型变量，表示该信号量的值，它与相应资源的使用情况有关；另一个是指向PCB的指针。当多个进程都等待同一信号量时，它们就排成一个队列，由信号量的指针项指出该队列的队首。（2分）

信号量通常可以简单反映出相应资源的使用情况，它与P、V操作原语一起使用可实现进程的同步和互斥。（1分）

P,V操作原语有如下定义。

P(S)顺序执行下述两个动作（1分）：

⑴信号量的值减1，即S=S-1;

⑵如果S>=0,则该进程继续执行。

如果S<0,则把该进程的状态置为阻塞态，把相应的PCB连入该信号量队列的末尾，并放弃处理机，进行等待（直到其他进程在S上执行V操作，把它释放出来为止）。

V(S)顺序执行下述两个动作（1分）：

⑴S值加1，即S=S+1;

⑵如果S>0,则该进程继续运行；

如果S<=0,则释放信号量队列上的第一个PCB所对应的进程（把阻塞态改为就绪态）,执行V操作的进程继续运行。

**9什么是线程？它与进程有什么关系？**

答：线程是进程中实施调度和分派的基本单位。

线程和进程之间有如下关系：

① 一个进程可以有多个线程，但至少有一个线程；而一个线程只能在一个进程的地址空间内活动。

② 资源分配给进程，同一进程的所有线程共享该进程的所有资源。

③ 处理机分给线程，即真正在处理机上运行的是线程。

④ 线程在执行过程中，需要协作同步。不同进程的线程间要利用消息通信的办法实现同步。

**10什么是管程？它由哪几部分组成？有什么基本特性?**

答：一个管程定义了一个数据结构和能为并发进程在其上执行的一组操作，这组操作能同步进程和改变管程中的数据。

一个管程由四个部分组成，它们是管程名称、局部与管程的共享数据的说明、对数据进行操作的一组过程和对该共享数据赋初值的语句。

管程具有以下特性：

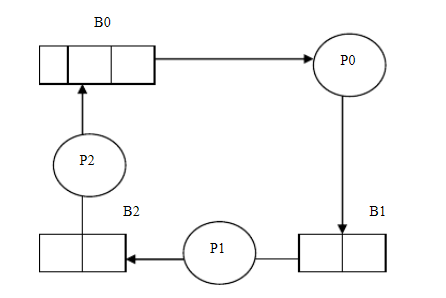
① 管程内部的局部数据变量只能被管程内定义的过程所访问，不能被管程外面声明的过程直接访问

② 进程要想进入管程，必须调用管程内的某个过程

③ 一次只能有一个进程在管程内执行，而其余调用该管程的进程都被挂起，等待该管程成为可用的。就是说，管程自身能有效地实现互斥。

**综合题**

**1如下图所示的工作模型中，有三个进程p0,p1,p2和三个缓冲区B0,B1,B2. 进程之间借助于相邻缓冲区进行消息传递：每个进程每次从缓冲区中取一条消息，经加工处理后送入另一个缓冲区中，三个缓冲区分别可存放3,2,2个消息。初始时，仅缓冲区0有一个消息。试用P、V操作写出三个进程之间的同步及互斥流程。**



答：这是一个生产者/消费者问题，而且每个进程既是生产者，也是消费者。(2’)

为此，应设置6个信号量：B0S1,B0S2,B1S1,B1S2,B2S1,B2S2,分别代表B0,B1,B2中是否有空缓冲和有数据。

B0S1,B0S2,B1S1,B1S2,B2S2:semaphore;

B0S1=2;B0S2=1;B1S1=2;B1S2=0;B2S1=2;B2S2=0;  (2’)

Cobegin  (`6’=2’\*3)

P0 P1 P2

begin begin begin

P(B0S2) P(B1S2) P(B2S2)

从B0取一个数据 从B1取一个数据 从B2取一个数据

V(B0S2) V(B1S1) V(B2S1)

加工 加工 加工

P(B1S1) P(B2S1) P(B0S1)

将加工结果送B1 将加工结果送B2 将加工结果送B0

V(B1S2) V(B2S2) V(B0S2)

end end end

coend

这道题也可以增加互斥信号量，以便P0与P1之间互斥使用B0缓冲区，P1与P2之间互斥使用B1缓冲区，P2与P0之间互斥使用B0缓冲区。这里主要描述它们之间的同步关系。若考虑互斥共享缓冲区，请自己加上。

**2设用三个队列管理缓冲区池的使用情况，分别为空白缓冲队列em，输入缓冲队列in，以及输出缓冲队列out。过程add\_buf(type,numb)和take\_buf(type,numb)分别用来把缓冲区numb插入type队列和从type队列中取出缓冲区numb。试描述进程从任一缓冲队列中得到一个缓冲区的过程get\_buf(type,numb)和释放一个缓冲区numb进入缓冲队列的过程put\_buf(type,numb)。**

答：假定用信号量s代表任一队列的可用缓冲区个数。假定三个队列的初值分别为n1,n2,n3。对任一队列的操作必须互斥。因此再引入一个互斥使用任一队列的信号量mutex，其初值为1。这里type代表队列的类型，它的取值为输入、输出和空白。(4’)

当有进程希望从任一队列取一个缓冲区时，过程get\_buf(type,numb)的动作如下：

get\_buf(type,numb)   (`3’)

begin

p(s)

p(mutex)

numb=take\_buf(type,numb)

v(mutex)

end

当有进程希望向任一队列送一个缓冲区时，过程put\_buf(type,numb)的动作如下：

put\_buf(type,numb)   (`3’)

begin

p(mutex)

add\_buf(type,numb)

v(mutex)

v(s)

end.

**3设有一个售票厅，可容纳100人购票。如果厅内不足100人则允许进入，进入后购票，购票后退出。如果厅内已有100人，则在厅外等候。试问：**

**1） 购票者之间是同步还是互斥？**

**用P、V操作表达购票者的工作过程。**

解：1）购票者之间是互斥关系。(2’)

2） 一个售票厅可容纳100人购票，说明最多允许100个购票者共享售票厅；可引入一个信号量empty，其初值为100。由于购票者必须互斥地进行购票，故应再设一个mutex，其初值为1。(4’)

用P、V操作表达购票者的工作过程如下：(`4’)

empty,mutex:semaphore;

empty:=100; mutex:=1;

begin

p(empty)

p(mutex)

进入厅内购票,购票后退出

v(empty)

v(mutex)

end.

**4某招待所有100个床位，住宿者入住要先登记（在登记表上填写姓名和床位号）．离去时要注销登记（在登记表上删去姓名和床位号）．请给出住宿登记及注销过程的算法描述．**

答：某招待所有100个床位，为了正确管理，引入一个信号量empty代表空床位数，初值为100；住宿者入住要先登记（在登记表上填写姓名和床位号），显然，登记表是一个临界资源，必须互斥访问，引入一个mutex，其初值为1。(4’)

住宿登记及注销过程的算法描述如下：

住宿登记：(`3’)

begin

p(empty)  //检查有无床位

p(mutex)  //申请登记

找出一个空床位将名字登入表中

v(mutex)

end

注销过程：(`3’)

begin

p(mutex)   //申请退房

找出自己的登记项，并删除该项的登记

v(mutex)

v(empty)

end.

**5有一个阅览室，共有100个座位。为了很好地利用它，读者进入时必须先在登记表上进行登记。该表表目设有座位号和读者姓名；离开时再将其登记项擦除。**

**试问：为描述读者的动作，应编写几个程序，应设几个进程、它们之间的关系怎样？并请用P、V操作描述进程之间的同步算法。**

解：为了描述阅览室，用一个登记表来记录其使用情况。表中共有100项。每当有读者进入阅览室时，为了正确地登记，各读者应互斥使用(1’)。为此设两个信号量：mutex为互斥信号量，用来制约各读者互斥地进行登记，其初值为1；empty为同步信号量，用来制约各读者能同时进入阅览室的数量，其初值为100  (2’)。

下面用两个过程描述对表格应执行的动作：

登记过程：(`2’) 擦除过程：(`2’)

begin begin

P(empty) P(mutex)

P(mutex) 找到自己的登记项擦除

找到一个登记项登记 V(mutex)

V(mutex) V(empty)

end end

为了正确地描述读者的动作，可以将读者看成进程。若干读者希望进入阅览室时，调用登记过程，退出阅览室时，调用擦除过程(1’)。

可见，一个程序可对应多个读者。可设的进程数由读者数决定，其动作如下：(`2’)

begin

调用登记过程

进入阅览室阅读

准备退出

调用擦除过程

end.

**6一条河上架设了由若干个桥墩组成的一座桥。若一个桥墩只能站一个人，过河的人只能沿着桥向前走而不能向后退。过河时，只要对岸无人过，就可以过；但不允许河对岸的两个人同时过，以防止出现死锁。请给出两个方向的人顺利过河的同步算法。**

解：假设一座桥由N个桥墩，也即最多允许有N个人同向过河，用一个计数器R记录同时过河的人数(2’)。用S1信号量保护计数器，其初值为1，R的初值为0；互斥使用桥的信号量用S表示，其初值为1。(2’)

同步算法描述如下：

procedure goriver()

begin

L:P(S1); //为同时过河,申请对计数器计数

  If R>N begin V(S1); goto L; end  //同方向过河的人站满桥墩时,重新申请计数

  R=R+1;

  If R==1 P(S); //申请过河

V(S1);   //释放计数器的使用权           (3’)

占有一个桥墩,并顺序过河到对岸;

P(S1);

  R=R-1;

  If R==0 V(S); //如果已经无同向的人过河,释放占用权

V(S1);                                                       (3’)

end.

**7在一个飞机订票系统中，多个用户共享一个数据库。各用户可以同时查询信息，若有一个用户要订票，须更新数据库时，其余所有用户都不可以访问数据库。请用P,V操作设计一个同步算法，实现用户查询与订票功能。要求：当一个用户订票而需要更新数据库时，不能因不断有查询者到来而使其长时间等待。利用信号量机制保证其正常执行。**

解：这是典型的读者——写者问题，查询信息的用户是读者，订票用户是写者，并且要求写者优先。(2’)

变量说明：(`2’)

计数变量

rc——正在运行的查询者进程数目，初值为0.

信号量

Sw——控制订票者进程的活动，初值为1.

Src——互斥使用rc变量，初值为1.

S——当订票者到达时封锁后续的读进程，初值为1.

读者进程

P(S)

P(Src)

rc=rc+1

if (rc==1) P(Sw)

V(Src)

V(S)             (2’)

查询库当中的信息

P(Src)

rc=rc-1;

if (rc==0) V(Sw)

V(Src)           (2’)

写者进程  (`2’)

P(S)

P(Sw)

更新数据库内容

V(Sw)

V(S)

**8某车站售票厅，任何时刻最多可容纳20名购票者进入，当售票厅中少于20名购票者时，则厅外的购票者可立即进入，否则需在外面等待。若把一个购票者看作一个进程，请回答下列问题：**

**(1)用PV操作管理这些并发进程时，应怎样定义信号量，写出信号量的初值以及信号量各种取值的含义。**

**(2)根据所定义的信号量，把应执行的PV操作填入下述空格中，以保证进程能够正确地并发执行。**

COBEGIN PROCESS PI(I=1，2，……)

begin

进入售票厅；

购票；

退出；

end

COEND

(3)若欲购票者最多为n个人，写出信号量可能的变化范围(最大值和最小值)。

答：(1)定义一信号量S，初始值为20。 (1’)

意义：(`3’=1’\*3)

S>0 S的值表示可继续进入售票厅的人数

S=0 表示售票厅中已有20名顾客(购票者)

S<0 |S|的值为等待进入售票厅的人数

(2)上空格为P(S)  (2’) ；下空格为V(S) (2’)

(3)S的最大值为20  (1’ )；S的最小值为20－n (1’ )

**9在公共汽车上，司机和售票员各行其职，司机负责开车和到站停车；售票员负责售票和开门关门，当售票员关好车门后，驾驶员才能开车行使。试用P/V操作实现司机与售票员间的同步。**

解答：semaphore mutex1=0,mutex2=0;  (2’)

  main(){

    cobegin

        driver()

        busman()

coend

}            (2’)

driver(){

    while(true){

    p(mutex1)

    启动公共汽车

    正常开车

    到站停车

    v(mutex2)

}

}           (3’)

busman(){

    while(true){

    关车门

    v(mutex1)

    售票

    p(mutex2)

    开车门

    上下乘客

}

}            (3’)

**10并发问题：设有两个优先级相同的进程p1, p2如下。令信号s1, s2的初值为0，已知z=2，试问p1, p2并发运行结束后x=? y=? z=?**

   进程p1              进程p2

   y := 1                x := 1

   y := y+2              x := x+1

   v(s1)                 p(s1)

   z := y+1              x := x+y

   p(s2)                 v(s2)

   y := z+y              z := x+z

解答：（分析过程略 2’）从结果来看，两个进程无论谁先谁后，结果都是一样的。(2’)

    x = 5;  y = 12;  z = 9  (6’)

**11设系统有三个并发进程R,C,P，共享一个能存放n个数据的环形缓冲区buf。进程R负责从输入设备上读数据，每读一个后把它存放在缓冲区buf的一个单元中；进程C负责从缓冲区读数据并进行处理，之后将处理结果再送入缓冲区的一个单元中；进程P负责从缓冲区读进程C处理的结果并打印。请用P、V操作为三进程的正确执行写出同步算法。**

解答：解决同步问题需设一个互斥信号量mux，用于控制三个进程互斥使用缓冲区，初值为1；再设三个同步信号量，用于控制对缓冲区的空闲数量和不同数据个数的记录。S0表示缓冲区空闲个数，初值为n；S1表示缓冲区中输入数据的个数，初值为0；S2表示缓冲区中输出数据的个数，初值为0。(4’)

算法描述如下：(`6’=2’\*3)

  进程R            进程C              进程P

  L1:               L2:                 L3:

P(S0)              P(S1)              P(S2)

P(mux)             P(mux)             P(mux)

读一个数据         从缓冲区中取一个   从缓冲区中读

送缓冲区           数据处理后放回去   输出数据

V(mux)             V(mux)            V(mux)

V(S1)              V(S2)              V(S0)  打印

gotoL1:            gotoL2:             gotoL3:

## 第三章  死锁

### 名词解释

**1死锁**

是指在一个进程集合中的每个进程都在等待仅由该集合中的另一个进程才能引发的事件而无限期地僵持下去的局面。

**2饥饿**

在系统中，每个资源占有者都在有限时间内释放它所占有的资源，但资源中存在某些申请者由于某种原因却永远得不到资源的一种错误现象。

**3死锁防止**

要求进程申请资源时遵循某种协议，从而打破产生死锁的四个必要条件中的一个或几个，保证系统不会进入死锁状态。

**4死锁避免**

对进程所发出的每一个申请资源命令加以动态地检查，并根据检查结果决定是否进行资源分配。就是说，在资源分配过程中若预测有发生死锁的可能性，则加以避免。这种方法的关键是确定资源分配的安全性。

**5安全序列**

针对当前分配状态来说，系统至少能够按照某种次序为每个进程分配资源（直至最大需求），并且使他们依次成功地运行完毕，这种进程序列{p1,p2,…,pn}就是安全序列。

**简答题**

**1计算机系统中产生死锁的根本原因是什么?死锁发生的四个基本条件是什么？**

答: 计算机系统中产生死锁的根本原因是：资源有限且操作不当 。死锁发生的四个基本条件有互斥条件、请求保持条件（占有且等待条件）、非剥夺条件（不可抢占条件）和环路条件（循环等待条件） 。

**2简述发生死锁的四个必要条件?**

答： 四个必要条件是：互斥条件、占有且等待条件（请求保持条件）、不可抢占条件（非剥夺条件）和循环等待条件（环路条件）。

互斥条件——某个资源在一段时间内只能由一个进程占有，不能同时被两个及其以上的进程占有。

占有且等待条件——进程至少已经占有一个资源，但又申请新的资源。

不可抢占条件——一个进程所占有的资源再用完之前，其他进程不能强行夺走资源，只能由该进程用完之后主动释放。

循环等待条件——存在一个进程等待序列{P1,P2,…,Pn}，其中，P1等待P2所占有的某个资源，P2等待P3所占有的某个资源，……，而Pn等待P1所占有的某个资源，从而形成一个进程循环等待。

**3什么是死锁？解决死锁的方法一般有那几种?**

答: 死锁是指在一个进程集合中的每个进程都在等待仅由该集合中的另一个进程才能引发的事件而无限期地僵持下去的局面。

解决死锁问题的一般方法为：死锁的预防、死锁的避免、死锁的检测和恢复。

**4死锁预防的基本思想是什么?死锁避免的基本思想是什么?**

答:死锁预防的基本思想是:要求进程申请资源是遵循某种协议,从而打破产生思索的四个必要条件中的一个或几个,保证系统不会进入死锁状态.

死锁避免的基本思想是:对进程所发出的每一个申请资源命令加以动态地检查,并根据检查结果决定是否进行资源分配.就是说,在资源分配过程中若预测有发生死锁的可能性,则加以避免.这种方法的关键是确定资源分配的安全性.

**5什么是死锁的安全序列?何谓系统是安全的？**

答：进程的安全序列{P1,P2,…,PN}是这样组成的：若对于每个进程Pi（1<=I<=n），它需要的附加资源可以被系统中当前可用资源加上所有进程Pj（j<i）当前占有资源之和所满足，则{ P1,P2,…,PN }为一个安全序列。

“系统是安全的”是指系统中的所有进程能够按照某种次序分配资源，并且依次运行完毕。即系统中的进程处于安全序列中。

**6资源按序分配法为什么能够预防死锁？**

证明：采用反证法来证明。

若存在循环等待，设在环路上的一组进程为{P0,P1,P2,…,Pn}，这里Pi等待进程Pi+1占有资源Ri（下角标取模运算，从而，Pn等待p0占有的资源）。由于Pi+1占有资源Ri,又申请资源Ri+1,从而一定存在F(i)<F(i+1), 该式对所有的i都成立。于是就有：

        F(R0)<F(R1)<…<F(Rn)<F(R0)

由传递性得到：

        F(R0)<F(R0)

显然，这是不可能的，因而，上述假设不成立，表明不会出现循环等待条件。

**7死锁和“饥饿”之间的主要差别是什么?**

答：死锁：多个并发进程相互等待对方占用的资源而产生的错误现象。

饿死：在系统中，由于系统采用的资源分配算法不当，虽然每个资源占有者都在有限时间内释放它所占的资源，但仍然使一些进程永远得不到资源的一种错误现象。

**综合题**

**1设系统中有三种类型的资源（A,B,C）和五个进程（P1,P2,P3,P4,P5）,A资源的数量为17，B资源的数量为5，C资源的数量为20。在T0时刻系统状态如表3-9所试。系统采用银行家算法来避免死锁。**

**①T0时刻是否为安全状态？若试，请给出安全序列。**

**②在T0时刻，若进程P2请求资源（0，3，4），能否实现资源分配？为什么？**

**③在②的基础上，若进程P4请求资源（2，0，1），能否实现资源分配？为什么?**

**④在③的基础上，若进程P1请求资源（0，2，0），能否实现资源分配？为什么?**

**表3-9 T0时刻系统状态**

**进程 最大资源需求量  已分配资源数量 系统剩余资源数量**

**A     B     C  A     B     C  A     B     C**

**P1  5     5      9  2     1      2  2     3      3**

**P2  5     3      6  4     0      2**

**P3  4     0     11  4     0      5**

**P4  4     2      5  2     0      4**

**P5  4     2      4  3     1      4**

解：

①T0时刻是安全状态，因为存在一个安全序列{P4,P5,P1,P2,P3}  (2’)

②不能实现资源分配，因为所剩余的资源数量不够。  (2’)

③可以分配。当分配完成后，系统剩余的资源向量为（0，3，2），这时，仍可找到一个安全序列{P4,P5,P1,P2,P3}  (3’)

④不能分配。如果分配的话，则系统剩余的资源向量为（0，1，2），这时无法找到一个安全序列。(3’)

**2在银行家算法中，系统有5个进程和3个资源。若出现以下资源分配情况：**

**进程 资源最大请求 已分配资源**

**p0 7, 5, 3 0, 1, 0**

**p1 3, 2, 2 2, 1, 0**

**p2 9, 0, 2 3, 0, 2**

**p3 2, 2, 2 2, 1, 1**

**p4 4, 3, 3 0, 0, 2**

**系统剩余资源数量为(3,2,2)。**

**1） 该状态是否安全(给出详细的检查过程)?**

**2） 如果进程依次有如下资源请求**

**p1:资源请求Request(1,0,2)?**

**p4:资源请求Request(3,3,0)?**

**p0:资源请求Request(0,1,0)?**

**则系统如何进行资源分配，才能避免死锁？**

解:

1)该系统状态是否安全,主要看能否找到一个进程完成序列.若能找到,系统只要按照这个序列为进程分配资源,所有进程就都可顺利完成;若找不到,系统状态就是不安全的.为此,可先求出进程的剩余请求矩阵.

进程 资源最大需求 已分配资源 剩余资源请求

P0 7, 5, 3 0, 1, 0 7, 4, 3

P1 3, 2, 2 2, 1, 0 1, 1, 2

P2 9, 0, 2 3, 0, 2 6, 0, 0

P3 2, 2, 2 2, 1, 1 0, 1, 1

P4 4, 3, 3 0, 0, 2 4, 3, 1

系统剩余资源向量A=(3,2,2)，在进程剩余资源请求矩阵中找，是否有一行，其值都小于或等于A.若有，选进程P1，满足它的全部资源请求，它在有限时间内能释放全部资源，并标记它为完成使系统剩余资源向量A=(5,3,2).之后再重复上述过程，从而找到了一个进城完成序列为：P1,P3,P4,P2,P0   (2’)。由此可见，系统状态是安全的(2’)。

2）p1:资源请求Request(1,0,2)时，由1）可知，可以立即满足它，使得A=(2,2,0),P1的分配向量为(3,1,2)，其剩余向量变为(0,1,0).  (2’)

p4:资源请求Request(3,3,0)时，由于系统剩余资源向量A=(2,2,0)，显然不能满足它的请求，因为系统剩余资源向量A小于P4的请求  (2’)

p0:资源请求Request(0,1,0)时，由于系统剩余资源向量A=(2,2,0)，若满足它的请求，使得系统剩余资源向量A=(2,1,0)。之后，系统仍可以找到一个进程完成序列P1,P4,P0,P4,P2。故可以满足它的请求。  (2’)

**3系统有同类资源10个，进程p1、p2和p3需要该类资源的最大数量分别为8，6，7。它们使用资源的次序和数量如下图所示。**

**1） 试给出采用银行家算法分配资源时，进行第5次分配后各进程的状态及各进程占用资源情况。**

**2） 在以后的申请中，那次的申请可以得到最先满足?给出一个进程完成序列。**

**次序 进程 申请量 次序 进程 申请量**

**1 P1 3 5 P2 2**

**2 P2 2 6 P1 3**

**3 P3 4 7 P3 3**

**4 P1 2 8 P2 2**

解：1）计算第5次分配后进程的状态和占用资源情况：(`5’=1’\*5)

① p1申请3个，满足，系统还剩7个

②p2申请2个，满足（因为系统的7个可以使p2运行完），系统还剩5个

③p3申请4个，因为若满足它的请求，可能使以后的任何进程都不能运行完，故p3等待

④p1申请2个，满足（系统还剩5个可以满足p1的最大请求），系统还剩3个

⑤ p2申请2个，不能满足，等待。此时系统的分配情况如下：

p1分配5个后正在运行，p2分配2个后等待分配2个，p3等待分配4个，系统还剩3个。

2）p1接着运行，p1申请3个可满足(2’)。P1运行完成后，释放资源，使系统的资源数量变为8个。首先将p3唤醒，满足它的4个资源，系统还剩4个，可以唤醒p2，满足它的2个请求。系统还剩2个。

P3申请3个，不能满足，等待。

P2申请2个，系统满足它，p2接着运行；p2完成，释放资源，使系统资源变为6个。系统唤醒p3，满足它的资源请求，最终p3完成，释放资源，使资源数量恢复为10个。

找到的进程完成序列为p1,p2,p3。 (3’)

**4设系统中有150个可用的同类资源。在某时刻系统中的进程已获得的资源和最大请求资源如下所示，请用银行家算法分别判断完成下列请求时，系统是否安全？若安全，请给出进程的完成序列。如不安全，请说明原因。**

**进程     最大需求量     当前已分配量**

**p1          70              25**

**p2          60              40**

**p3          60              45**

**p4          60               0**

**(1) 进程p4当前请求25个资源；**

**(2) 之后p4又提出35个资源的请求。**

解答：系统当前剩余资源量为：150 – 25 – 40 – 45 = 40  (2’)

(1) 可以满足(2’)，假定先分配p4的25个资源，系统还剩15个。将这15个资源可先分配给p3，p3达到最大请求，释放60个；之后可以分配给其他任何进程，系统中的进程都能顺利完成。由此可见，p2请求的25个资源可以满足，且能找到完成序列：p3，p1，p2，p4，…(4’)

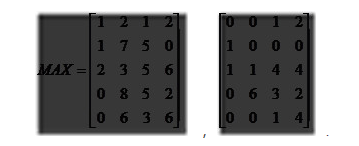
(2) 当p4再提出35个资源请求时，系统还剩15，显然不能满足它的请求，让其阻塞等待。(2’)

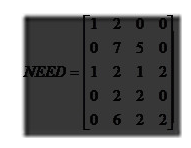
5系统中有五个进程，分别为p1\p2\p3\p4\p5，四类资源分别为r1\r2\r3\r4。某一时刻，系统剩余资源向量A=(1，2，3，0)。

（1） 用银行家算法试判断系统当前状态是否安全？

（2） 当进程p3提出对资源r3的剩余请求时，能否满足她？

（3） 系统初始配置的各类资源分别为多少？

  .

解答：系统剩余资源向量 A=(1, 2, 3, 0) 。现在需求出各进程的剩余资源请求矩阵:

(2’)

(1) 详细步骤省略。由于系统存在一个进程完成的安全序列P1\P3\P4\P2\P5(2’)，故系统状态是安全的(2’)。

(2) 进程P3提出对资源R3的剩余请求为1，由于系统剩余资源向量A=(1,2, 3, 0)，故可以假定分配给它。如果能找到一个安全序列，就可以真正进行分配。当分配给P3一个资源时，系统剩余资源向量A=(1 ,2 ,2 , 0)。由此可见，仍然可以找到一个与(1)相同的安全序列。故可以满足P3的请求。(3’)

(3) 系统初始配置的各类资源分别为(3 ,9 , 12 , 12 )。(1’)

## 第四章  调度

### 名词解释

**1作业**

用户在一次上机过程中要求计算机系统所做工作的集合。

**2周转时间**

是指从作业进入系统开始，到作业退出系统所经历的时间。

**3响应时间**

是分时系统的一个技术指标，指从用户输入命令到系统对命令开始执行和显示所需要的时间。

**4作业调度**

作业调度的主要任务是完成作业从后备状态到执行状态和从执行状态到完成状态的转换。

**5进程调度**

也称低级调度程序，它完成进程从就绪状态到运行状态的转化。实际上，进程调度完成一台物理的cpu转变成多台虚拟（或逻辑）的cpu的工作。

**6交换调度**

是基于系统确定的某个策略，将主存中处于等待状态或就绪状态的某个或某些进程交换到外存交换区中，以便将外存交换区上具备运行条件的进程换入主存，准备执行。引入交换调度的目的是为了解决主存紧张和提高主存的利用效率。

**7剥夺式调度**

当一个进程正在执行时，系统基于某种策略强行将处理机从占有者进程剥夺而分配给另一个进程的调度。这种调度方式系统开销大，但系统能及时响应请求。

**8非剥夺式调度**

系统一旦把处理机分配给某个进程之后，该进程一直运行下去，直到该进程完成或因等待某个事件发生时，才将处理机分配给其他进程。这种调度方式实现简单，系统开销小，但系统性能不够好。

**简答题**

**1作业由哪几部分组成？各有什么功能？**

答：作业由三部分组成：程序、数据和作业说明书。

程序和数据完成用户所要求的业务处理工作，作业说明书则体现用户的控制意图。

**2试比较作业和进程的区别**

答：一个进程是一个程序对某个数据集的执行过程，是分配资源的单位。作业是用户需要计算机完成某项任务，而要求计算机所做工作的集合。一个作业的完成要经过作业提交、作业收容、作业执行和作业完成4个阶段。而进程是已提交完毕的程序所执行过程的描述，是资源分配的基本单位。

其主要区别关系如下：

（1）作业是用户向计算机提交任务的任务实体。在用户向计算机提交作业之后，系统将它放入外存中的作业等待队列中等待执行。而进程则是完成用户任务的执行实体，是向系统申请分配资源的基本单位。任一进程，只要它被创建，总有相应的部分存在内存中。

（2）一个作业可由多个进程组成。且必须至少由一个进城组成，但反过来不成立。

（3）作业的概念主要用在批处理系统中。像UNIX这样的分时系统中，则没有作业概念。则进程的概念则用在几乎所有的多道程序系统中。

**3高级调度与低级调度的主要功能是什么？为什么要引入中级调度？**

答:高级调度的主要功能是根据一定的算法,从输入的一批作业中选出若干作业,分配必要的资源,如内存、外设等，为它建立相应的用户作业进程和为其服务的系统进程（如输入/输出进程），最后把它们的程序和数据调入内存，等待进程调度程序对其执行调度，并在作业完成后做善后处理工作。

低级调度的主要功能是根据一定的算法将cpu分派给就绪队列中的一个进程。

为了使内存中同时存放的进程数目不至于太多，有时需要把某些进程从内存移到外存上，以减少多道程序的数目，为此设立了中级调度.

**4处理机调度一般分为哪三级？其中哪一级调度必不可少？为什么？**

答：处理机调度一般可分为高级调度（作业调度）、中级调度和低级调度（进程调度） 。其中进程调度必不可少 。

    进程只有在得到CPU之后才能真正活动起来，所有就绪进程经由进程调度才能获得CPU的控制权。实际上，进程调度完成一台物理的CPU转变成多台虚拟机（或逻辑）的CPU的工作，进程调度的实现策略往往决定了操作系统的类型，其算法优劣直接影响整个系统的性能。

**5作业调度与进程调度之间有什么差别？二者间如何协调工作？**

答：作业调度与进程调度之间的差别主要是：作业调度是宏观调度，它所选择的作业只是具有获得处理机的资格，但尚未占有处理机，不能立即在其上实际运行；而进程调度是微观调度，动态地把处理机实际地分配给所选择的进程，使之真正活动起来。另外，进程调度相当频繁，而作业调度执行的次数一般很少。

作业调度从外存的后背队列中选择一批作业调入内存，为它们创建进程，这些进程被送入就绪队列。进程调度从就绪队列中选出一个进程来，并把它的状态改为运行态，把cpu分配给它。当运行进程要等待某一事件时，就让出cpu，进入相应的阻塞队列，并进行进程调度。运行进程完成后，由作业调度进行善后处理工作。

**综合题**

**1有A、B、C、D、E，共5个待运行作业，各自估计的运行时间为9，6，3，5，x。试问采用哪种运行次序使得平均响应时间为最短？（答案依赖于x）**

解答：

由于短作业优先调度算法可以使作业的平均周转时间最短，同样使作业的平均响应时间为最短。  (5’)

  下面对x的取值进行讨论：(`5’=1’\*5)

  当0<x<=3时，作业的运行顺序应为E(x),C(3),D(5),B(6),A(9);

  当3<x<5时，作业的运行顺序应为C(3),E(x),D(5),B(6),A(9);

  当5<=x<=6时，作业的运行顺序应为C(3),D(5),E(x),B(6),A(9);

  当6<x<=9时，作业的运行顺序应为C(3),D(5),B(6),E(x),A(9);

  当x>9，作业的运行顺序应为C(3),D(5),B(6),A(9),E(x)

**4有一个具有如下作业流的批处理处理系统，作业调度采用短作业优先，进程调度采用基于优先数的抢先式调度算法。下表给出的是作业序列和相应进程的优先数，优先数越小优先级越高。**

作业名    到达时间   估计运行时间/min   优先数

   1          8:00          40              4

   2          8:20          30              2

   3          8:30          50              3

   4          8:50          20              5

(1) 列出所有作业进入内存时间及完成时间

(2) 计算作业的平均周转时间和平均带权周转时间

解答：

(1)作业进入内存时间与结束时间如下所示：(`4’=1’\*4)

作业名   进入内存时间   结束时间

   1         8:00          9:10

   2         8:20          8:50

   3         9:10          10:00

   4         8:50          10:20

(2)各作业的周转时间为：   (`4’=1’\*4)

作业A：9:10 – 8:00 = 70 min

作业B：8:50 – 8:20 = 30 min

作业C：10:00 – 8:30 = 90 min

作业D：10:20 – 8:50 = 90 min

作业的平均周转时间为：(70+30+90+90)/4=70 min    (1’)

作业的平均带权周转时间为：(70/40+30/30+90/50+90/20)/4=2.26 min   (1’)

## 第五章 存储管理

### 名词解释

**1物理地址**

内存中各存储单元的地址由统一的基地址顺序编址，这种地址称为物理地址。

**2逻辑地址**

用户程序经编译之后的每个目标模块都以0为基地址顺序编址，这种地址称为逻辑地址。

**3逻辑地址空间**

由程序中逻辑地址组成的地址范围叫做逻辑地址空间。

**4物理地址空间**

由内存中的一系列存储单元所限定的地址范围称作内存空间。

**5重定位**

把逻辑地址转变为内存物理地址的过程叫做重定位。

**6静态重定位**

在目标程序装入内存时所进行的重定位。

**7动态重定位**

在程序执行期间，每次访问内存之前进行的重定位。

**8内部碎片**

在一个分区内部出现的碎片（即被浪费的空间）称作内部碎片。如固定分区法会产生内部碎片。

**9外部碎片**

在所有分区之外新产生的碎片称作外部碎片，如在动态分区法实施过程中出现的越来越多的小空闲块，由于它们太小，无法装入一个小进程，因而被浪费掉。

**10碎片**

在分区法中，内存出现许多容量太小、无法被利用的小分区称作“碎片”。

**11紧缩**

移动某些已分区的内容，使所有作业的分区紧挨在一起，而把空闲区留在另一端，这种技术称为紧缩。

**12可重定位地址**

当含有它的程序被重定位时，将随之被调整的一种地址。

**13固定分区法**

内存中分区的个数固定不变，各个分区的大小也固定不变，但不同分区的大小可以不同，每个分区只可装入一道作业。

**14动态分区法**

各个分区是在相应作业要求进入内存时才建立的，使其大小恰好适应作业的大小。

**15可再入代码**

也称纯代码，是指那些在其执行过程本身不做任何修改的代码，通常由指令和常数组成。

**16虚拟存储器**

虚拟存储器是用户能作为可编程内存对待的虚拟存储空间，在这种计算机系统中实现了用户逻辑存储器与物理存储器的分离，它是操作系统给用户提供的一个比真实内存空间大得多的地址空间。

**17抖动**

页面抖动是系统中频繁进行页面置换的现象。即如果一个进程没有一定数量的内存块，它很快就发生缺页。此时，它必须淘汰某页。由于所有这些页面都正在使用，所以刚被淘汰出去的页很快又被访问，因而要把它重新调入。可是调入不久又再被淘汰出去，这样再访问，再调入，如此反复，使得整个系统的页面替换非常频繁，以致大部分机器时间都用在来回进行的页面调度上，只有一小部分时间用于进程的实际运算方面。

**18工作集**

工作集是一个进程在某一小段时间内访问页面的集合。利用工作集模型可防止抖动，也可以进行页面置换。

**19程序局部性原理**

在相对短的一段时间内，进程集中在一组子程序或循环中之行，导致所有的存储器访问局限于进程地址空间的一个固定子集。这种现象就叫做程序局部性原理。

**20快表**

又叫“联想存储器”。在分页系统中，由于页表是存放在主存中的，因此cpu存取一个数据时要访问两次主存。这样使计算机的处理速度降低约一倍。为了提高地址变换速度，在地址变换机构中增设一个具有并行查找能力的高速缓冲存储器，用以存放当前访问的页表项。这样的高速缓冲存储器就是快表。

**21交换**

交换系统指系统根据需要把主存中暂时不运行的某个（或某些）作业部分或全部移到外存。而把外存中的某个（或某些）作业移到相应的主存区，并使其投入运行。

**22换页**

指系统根据某种策略选择某页出主存，将某页调入主存的过程。

**23实存**

实存是指计算机配置的物理存储器，它直接向cpu提供程序和数据。

**24虚存**

虚存是指系统向用户程序提供的编程空间，其大小由cpu的地址长度决定。

**简答题**

**1解释固定分区法和动态分区法的基本原理。**

答：固定分区法——内存中分区的个数固定不变，各个分区的大小也固定不变，但不同分区的大小可以不同。每个分区只可装入一道作业。

动态分区法——各个分区是在相应作业要进入内存时才建立的，使其大小恰好适应作业的大小。

**2说明内部碎片和外部碎片的不同之处**

答：内存中出现的其容量太小、无法被利用的小分区称作碎片 。内部碎片和外部碎片出现的位置不同 。内部碎片出现在一个分区的内部（即被浪费的空间），如固定分区法会产生内部碎片 。外部碎片出现在所有分区之外，是新增的小分区，如在动态分区法实施过程中会出现外部碎片 。

**3动态重定位分区管理方式中如何实现虚-实地址映射？**

答：作业装入内存时，是将该用户的程序和数据原封不动地装入到内存中 。当调度该进程在cpu上执行时，操作系统就自动将该进程在内存的起始地址装入基址寄存器，将进程的大小装入限长寄存器 。当执行指令时，如果地址合法，则将相对地址与基址寄存器中的地址相加，所得结果就是真正要访问的内存地址；如果地址越界，则发出相应中断，进行处理 。

**4什么是虚拟存储器？它有哪些基本特征？**

答:虚拟存储器是用户能作为可编址内存对待的虚拟存储空间,在这种计算机系统中实现了用户逻辑存储器与物理存储器的分离,它是操作系统给用户提供的一个比真实内存空间大得多的地址空间。

虚拟存储器的基本特征是:虚拟扩充——不是物理上,而是逻辑上扩充了内存容量;部分装入——每个作业不是全部一次性地装入内存,而是只装入一部分;离散分配——不必占用连续的内存空间,而是”见缝插针”;多次对换——所需的全部程序和数据要分成多次调入内存。

5引入虚拟存储器后，除了获得主存“扩充”的好处，还有什么好处？

答：引入虚存后，程序的地址空间都是虚地址的集合，只有在程序运行中通过硬件地址转换机构和操作系统的相应软件，才能将虚地址变换成主存的实地址，这将为主存的分配带来更大的灵活性。另外，虚、实地址分开，用户程序不能干扰实地址的生成，从而实现了存储器的保护 。

**6什么是分页？什么是分段？二者有何主要区别？**

答：分页是由系统将一个进程的逻辑地址空间划分成若干大小相等的部分，每一部分称做一个页面。

分段是用户根据作业的逻辑关系进行自然划分，每个分段是作业中相对独立的一部分。

分段和分页都是非连续的存储管理方法，

分页和分段的主要区别有：

①页是信息的物理单位，段是信息的逻辑单位。

②页面的大小由系统确定，并且各页大小都相同；各段长度因段而已，由用户决定。

③分页的作业地址空间是一维的，分段的作业的地址空间是二维的。

④分页的活动对用户是不可见的，而分段是用户可见的活动。

**7在分页系统中页面大小由谁决定？页表的作用是什么？如何将逻辑地址转换成物理地址？**

答:在分页系统中页面大小由硬件决定。

页表的作用是:实现从页号到物理块号的地址映射。

逻辑地址转换成物理地址的过程是:用页号P去检索页表，从页表中得到该页的物理块号，把它装入物理地址寄存器中。同时，将页内地址d直接送入物理地址寄存器的块内地址字段中。这样，物理地址寄存器中的内容就是由二者拼接成的实际访问内存地址，从而完成了从逻辑地址到物理地址的转换。

**8什么是belady现象？**

答：belady现象是指在使用FIFO算法进行内存页面置换时 ，在未给进程或作业分配足它所要求的全部页面的情况下，有时出现的分配的页面数增多，缺页次数发而增加的奇怪现象。

**9请求分页技术的基本思想是什么？它与简单分页技术之间有何根本区别？**

答：请求分页技术的基本思想是：当一个进程的部分页面在内存时就可调度它运行；在运行过程中若用到的页面尚未在内存，则把它们动态换入内存。这样，就减少了对换时间和所需内存数量，允许增加程序的道数。

请求分页技术是在简单分页技术基础上发展起来的，两者根本区别是：请求分页提供虚拟存储器，而简单分页系统并未提供虚拟存储器。

**10为什么分段技术比分页技术更容易实现程序或数据的共享和保护？**

答： 每一段在逻辑上是相对完整的一组信息，分段技术中的共享是在段一级出现的。这样，任何共享的信息就可以单独成为一段。同样，段中所有内容可以用相同的方式进行使用，从而规定相同的保护权限。 然而，页是信息的物理单位，在一页中可能存在逻辑上互相独立的两组或多组信息，各有不同的使用方式和存取权限，因而，对分页难以进行共享和保护。

**11何谓工作集？它有什么作用？**

答：工作集是一个进程在某一小段时间内访问页面的集合。

利用工作集模型可防止抖动，也可以进行页面置换。

**12什么是页面抖动？系统怎样检测是否出现抖动？一旦检测到抖动？系统如何消除它？**

答：页面抖动是系统频繁进行页面置换的现象。整个系统的页面替换非常频繁，以致大部分机器时间都用在来回进行的页面调度上，只有一小部分时间用于进程的实际运算方面。

操作系统监督每个进程的工作集，并给它分配工作集所需的内存块。若有足够多的额外块，就可以装入并启动另外的进程。如果工作集增大了，超出可用块的总数，即系统中全部进程对内存块的总请求量大于可用内存块的总量，将出现抖动，因为某些进程得不到足够的内存块。

一旦检测到抖动，操作系统要选择一个进程让它挂起，把它的页面写出去，把它占用的内存块分给别的进程。被挂起的进程将在以后适当时机重新开始执行。

**综合题**

**1考虑下面页面走向：1，2，3，4，2，1，5，6，2，1，2，3，7，6，3，2，1，2，3，6 当内存块数量分别为3时，试问LRU,FIFO,OPT三种置换算法的缺页次数各是多少？（注意，所有内存最初都是空的，凡第1次用到的页面都产生一次缺页）**

答：

LRU

1 2 3 4 2 1 5 6 2 1 2 3 7 6 3 2 1 2 3 6

1 1 1 4 4 4 5 5 5 1 1 1 7 7 7 2 2 2 2 2

2 2 2 2 1 1 1 2 2 2 2 2 6 6 6 1 1 1 6

3 3 3 3 3 6 6 6 6 3 3 3 3 3 3 3 3 3

× × × × × × × × × × × × × × ×

(2’)

FIFO

1 2 3 4 2 1 5 6 2 1 2 3 7 6 3 2 1 2 3 6

1 1 1 4 4 4 4 6 6 6 6 3 3 3 3 2 2 2 2 6

2 2 2 2 1 1 1 2 2 2 2 7 7 7 7 1 1 1 1

3 3 3 3 5 5 5 1 1 1 1 6 6 6 6 6 3 3

× × × × × × × × × × × × × × × ×

(2’)

OPT

1 2 3 4 2 1 5 6 2 1 2 3 7 6 3 2 1 2 3 6

1 1 1 1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3 3

2 2 2 2 2 2 2 2 2 2 2 7 7 7 2 2 2 2 2

3 4 4 4 5 6 6 6 6 6 6 6 6 6 1 1 1 6

× × × × × × × × × × × (2’)

内存块数 置换算法

FIFO LRU OPT

3 16 15 11        (3’)

**2考虑下面存储访问序列，该程序大小为460字：10，11，104，170，73，309，185，245，246，434，458，364 设页面大小是100字，请给出该访问序列的页面走向。又设该程序基本可用内存是200字，采用FIFO置换算法，求出缺页率。如果采用LRU算法，缺页率是多少？如果采用最优淘汰算法，其缺页率又是多少？**

解:

该序列的页面走向为：0、1、0、3、1、2、4、3。  (1’)

FIFO

0 1 0 3 1 2 4 3

0 0 0 3 3 3 4 2

1 1 1 1 2 2 3

× × × × × ×    (2’)

LRU

0 1 0 3 1 2 4 3

0 0 0 0 1 1 4 4

1 1 3 3 2 2 3

× × × × × × ×    (2’)

OPT

0 1 0 3 1 2 4 3

0 0 0 3 3 3 3 3

1 1 1 1 2 4 4

× × × × × (2’)

算法 FIFO LRU OPT

缺页次数 6 7 5

缺页率 6/12=0.5 7/12=0.583 5/12=0.417     (3’)

**3设某页系统中，页帧大小为100字。一个程序大小为1200字，可能的访问序列如下：**

**10，205，110，735，603，50，815，314，432，320，225，80，130，270**

**系统采用LRU算法。当为其分配4个主存块时，给出该作业驻留的各个页的变化情况及页故障数。**

答：

首先将逻辑地址变换成页号。这样10，205，110，735，603，50，815，314，432，320，225，80，130，720，通过除以页的大小100，页号分别为0，2，1，7，6，0，8，3，4，2，0，1，2。(3’)

系统为运行进程分配4个主存块，采用LRU算法，因此可以列表给出进程的缺页情况：

0 2 1 7 6 0 8 3 4 3 2 0 1 2

0 2 1 7 6 0 8 3 4 3 2 0 1 2

0 2 1 7 6 0 8 3 4 3 2 0 1

0 2 1 7 6 0 8 8 4 3 2 0

0 2 1 7 6 0 0 8 4 3 3

F F F F F F F F F S F F F S      (5’)

由上表可见，被淘汰的页依次为0，2，1，7，6，0，8，4。缺页次数为12次 (2’)

**4某请求页式管理系统，用户编程空间有40个页面，每个页面为200H字节。假定某时刻用户页表中虚页号和物理块号对照表如下：**

**虚页号 0 2 5 17 20**

**物理块号 5 20 8 14 36**

**求虚地址0A3CH、223CH分别对应的物理地址。**

答：

虚地址0A3CH转换成十进制数为2620，每个页为200H，即512B，由2620/512可得，页号为5，页内地址为60。查页表可知，其主存块号为8。(3’)

因此地址为2620的物理地址为：8\*512+60=4156。(2’)

虚地址223CH转换成十进制数为8762，由8762/512可得，其页号为17，页内地址为58。查页表可知，其主存块号为14。(3’)

因此地址为8762的物理地址为14\*512+58=7226。(2’)

**5某系统采用页式存储管理策略，拥有逻辑空间32页，每页2KB；拥有物理空间1MB。**

**1） 写出逻辑地址的格式**

**2） 若不考虑访问权限位，进程的页表有多少项？每项至少多少位？**

**3） 如果物理空间减少一半，页表结构应作怎样的改？**

答：

1）逻辑空间32页，占5个二进制位。每页2KB，占11位。故描述逻辑空间需要16位(2’)。

15 … 11 10 … 0

逻辑地址的格式：[       |                  ]   (1’)

2)进程的页表有32项，每项的位数由主存的分块数决定(2’)。1MB的空间可划分为512个2KB的块，每个块用9个二进制位表示(2’)。

3）如果物理空间减少一半时，主存地址需要19位表示，仍大于逻辑空间的大小，故页表结构可以不变。(3’)

**6有一虚拟存储系统，采用先进先出（FIFO）的页面淘汰算法。在主存忠为每一个作业进程开辟3页。某作业运行中使用的操作数所在的页号依次为：4，3，2，1，4，3，5，4，3，2，1，5。**

**1） 该作业运行中总共出现多少次缺页？**

**2） 若每个作业进程在主存拥有4页，又将产生多少次缺页？**

**3） 如何解释所出现的现象？**

解：

先进先出算法的实质是：总是选择作业中在主存驻留时间最长的一页进行淘汰。

若在主存中为每一作业进程开辟3页，对于题中的页面访问过程，其页面调度过程如下所示

4 3 2 1 4 3 5 4 3 2 1 5

页面1 4 4 4 1 1 1 5 5 5 5 5 5

页面2 3 3 3 4 4 4 4 4 2 2 2

页面3 2 2 2 3 3 3 3 3 1 1

缺页中断 F F F F F F F F F         (3’)

1） 该作业运行中总共出现9次缺页(1’)

2） 在主存拥有4页，又将产生10次缺页(1’)。其页面调度过程见下图：

4 3 2 1 4 3 5 4 3 2 1 5

页面1 4 4 4 4 4 4 5 5 5 5 1 1

页面2 3 3 3 3 3 3 4 4 4 4 5

页面3 2 2 2 2 2 2 3 3 3 3

页面4 1 1 1 1 1 1 2 2 2

缺页中断 F F F F F F F F F F        (3’)

3)从这个例子可以看出，当主存中为每一作业进程开辟4页时，出现了缺页次数反而增加的现象。这种现象称为Belady现象。(2’)

**7关于存储管理，试问：**

**（1） 在分页、分段和段页式存储管理中，当访问一条指令或数据时，需要访问内存几次？各做什么处理？**

**（2） 假设一个分页存储系统具有快表，多数活动页表都可以存在其中，页表放在内存中，内存访问时间是1us。若快表的命中率是85%，快表的访问时间为0.1us，则有效存取时间为多少？若快表命中率为50%，那么有效存取时间为多少？**

解答：（1）分页需要访问2次，第一次访问页表，第二次执行访内操作(2’)；分段需要访问2次，第一次访问段表，第二次执行访内操作；段页式需要访问3次，第一次访问段表，第二次访问页表，第三次执行访内操作(2’)。

（2）当快表的命中率为85%时，执行一次访内操作需要的时间：

   T=1\*0.85+2\*(1-0.85)=1.15(us)   (3’)

当快表的命中率为50%时，执行一次访内操作需要的时间：

   T=1\*0.5+2\*(1-0.5)=1.5(us)   (3’)

**8在一个采用页式虚拟存储管理的系统中，有一用户作业，它依次要访问的字地址序列是：115，228，120，88，446，102，321，432，260，167，若该作业的第0页已经装入主存，现分配给该作业的主存共300字，页的大小为100字，请回答下列问题：**

**（1）按FIFO调度算法将产生多少次缺页中断，依次淘汰的页号为多少，缺页中断率为多少。**

**（2）按LRU调度算法将产生多少次缺页中断，依次淘汰的页号为多少，缺页中断率为多少。**

答：

页面走向为：1，2，1，0，4，1，3，4，2，1

（1）按FIFO调度算法将产生5次缺页中断；依次淘汰的页号为：0，1，2； 缺页中断率为：5/10=50%    (3’)

1 2 1 0 4 1 3 4 2 1

0 0 0 0 0 4 4 4 4 4 4

1 1 1 1 1 1 3 3 3 3

2 2 2 2 2 2 2 2 1

× × × × ×      (2’)

（2）按LRU调度算法将产生6次缺页中断；依次淘汰的页号为：2，0，1，3； 缺页中断率为：6/10=60%   (3’)

1 2 1 0 4 1 3 4 2 1

0 0 0 0 0 0 0 3 3 3 3

1 1 1 1 1 1 1 1 2 2

2 2 2 4 4 4 4 4 1

× × × × × ×   (2’)

**9一台计算机含有65536字节的存储空间，这一空间被分成许多长度为4096字节的页。有一个程序，其代码段为32768字节，数据段为16386字节，栈段为15870字节。试问该机器的主存空间适合这个进程吗？如果将每页改成512字节，合适吗？**

答：

当存储空间每块为4096B时，共可分成16块。其中：

程序代码段占：32768/4096=8块；   (1’)

数据段占：16386/4096=5块；  (1’)

栈段占：15870/4096=4块；  (1’)

合计为：8+5+4=17块；  (1’)

故该机器的主存空间不适合这个作业。 (1’)

当存储空间每块为512B时，共可分成128块。其中：

程序代码段占：32768/512=64块； (1’)

数据段占：16386/512=32块； (1’)

栈段占：15870/512=31块； (1’)

故合计为：64+32+31=127块。 (1’)

故该机器的主存空间是适合这个作业的。 (1’)

**10一个请求分页系统中，内存的读/写周期为8纳秒，当配置有快表时，查快表需要1纳秒，内外存之间传送一个页面的平均时间为5000纳秒。假定快表的命中率为75%，页面失效率为10%，求内存的有效存取时间。**

答：

访问主存的时间可用下面公式表示：

访问主存时间=主存的命中率\*（快表的命中率\*访问快表的时间+执行实际操作访问主存的时间）+页面失效率\*页面失效时的访问时间   (6’)

因此 TA=(1-0.1)\*[0.75\*1+(1-0.75)\*(8+1)+8]+0.1\*5000   (4’)

**11在一个虚拟存储器系统中，一次访问主存的时间用TA1表示，一个访问外存的时间用TA2表示。假定TA1=10^-7秒，TA2=10^-2秒。试问，为了使访问效率达到80%以上，命中率H至少应该达到多少？**

答：

访问效率：e=TA1/TA=0.8   (2’)

TA=TA1/0.8=1.25\*10^-7 s    ( 2’)

TA=H\*TA1+(1-H)\*TA2=H(TA1-TA2)+TA2

H=(TA-TA2)/(TA1-TA2)  (2’)

解得  H=(1.25\*10^-7-10^-2)/(10^-7 – 10^-2)=0.999975  (2’)

因此，为了使访问效率达到80%以上，命中率H至少应该达到0.999975。(2’)

## 第六章  文件系统

### 名词解释

**1逻辑记录**

用户构造文件时使用的一个信息单位。通常以逻辑记录为单位存取文件。

**2物理记录**

文件存储器上组织信息的一个单位。它是文件存储器识别信息的单位。

**3文件**

是命名的相关信息的集合体，它通常存放在外存（如磁盘、磁带）上，可以作为一个独立单位存放并实施相应的操作（如打开、关闭、读、写等）。

**4文件系统**

操作系统中负责操纵和管理文件的一整套设施，它实现文件的共享和保护，方便用户“按名存取”。

**5目录项**

为了加快对文件的检索，把文件控制块集中在一起进行管理。这种文件控制块的有序集合称为文件目录。当然，文件控制块也是其中的目录项。

**6目录文件**

全由目录项构成的文件成为目录文件。

**7路径**

在树形目录结构中，从根目录出发经由所需子目录到达指定文件的通路。

**8当前目录**

为节省文件检索的时间，每个用户可以指定一个目录作为当前工作目录，以后访问文件时，就从这个目录开始向下顺序检索。这个目录就称作当前目录。

**9文件的逻辑组织**

用户对文件的观察和使用是从自身处理文件数据时所采用的组织方式来看待文件组织形式。这种从用户观点出发所见到的文件组织形式称为文件的逻辑组织。

**10文件的物理组织**

文件在存储设备上的存储组织形式称为文件的物理组织。

**11文件控制块**

用于描述和控制文件的数据结构，其中包括文件名、文件类型、位置、大小等信息。文件控制块与文件一一对应，即在文件系统内部，给每个文件唯一地设置一个文件控制块，核心利用这种结构对文件实施各种管理。

**12存取权限**

用户或系统为文件规定的谁能访问，以及如何访问的方式

**简答题**

**1什么是文件、文件系统？文件系统有哪些功能？**

答：在计算机系统中，文件被解释为一组赋名的相关字符流的集合，或者是相关记录的集合。

文件系统是操作系统中与管理文件有关的软件和数据。

文件系统的功能是为用户建立文件，撤销、读写修改和复制文件，以及完成对文件的按名存取和进行存取控制。

**2文件系统一般按什么分类？可以分为那几类？**

答：文件系统一般按性质，用途，组织形式，文件中的信息流向或文件的保护级别等分类：按文件的性质与用途可以分为系统文件，库文件和用户文件。按文件的组织形式可以分为普通文件，目录文件和特殊文件。按文件中的信息流向可以分为输入文件，输出文件和输入/输出文件。按文件的保护级别可以分为只读文件，读写文件，可执行文件和不保护文件。

**3什么是文件的逻辑结构，什么是记录？**

答：文件的逻辑结构就是用户可见的结构，可分为字符流式的无结构文件和记录式的有结构文件两大类。

记录是一个具有特定意义的信息单位，它由该记录在文件中的逻辑地址（相对位置）与记录名所对应的一组关键字，属性及其属性值所组成。

**4什么是文件目录？文件目录中包含那些信息？**

答：一个文件的文件名和对该文件实施控制管理的说明信息称为该文件的说明信息，又称为该文件的目录。

文件目录中包含文件名、与文件名相对应的文件内部标识以及文件信息在文件存储设备上第一个物理块的地址等信息。另外还可能包含关于文件逻辑结构、物理结构、存取控制和管理等信息。

**5文件系统中目录结构主要有哪几种？分别说明各自的实现思想？**

答：文件系统中的目录结构主要有：单级目录结构，二级目录结构，树形目录结构和非循环图目录结构。

单级目录结构——在这种组织方式下，全部文件都登记在同一目录中。

二级目录结构——在主文件目录中登载了各个用户的名称，每个用户有自己的用户文件目录。

树形目录结构——在这种结构中，只有一个根目录，每一级目录可以是下级目录的说明，也可以是包含文件的说明。从根开始一层一层地扩展下去，就形成一个树形层次结构。

非循环图目录结构——树形目录结构的自然推广就是非循环图目录结构，它允许一个文件或目录可在多个父目录中占有项目，但并不构成环路。

**6什么是文件控制块?它与文件有何关系？**

答：文件控制块——用于描述和控制文件的数据结构，其中包括文件名、文件类型、位置、大小等信息。

文件控制块与文件一一对应，即在文件系统内部，给每个文件唯一地设置一个文件控制块，核心利用这种结构对文件实施各种管理。

**7文件系统中的目录结构有哪几种基本形式？各有何优缺点？**

答：文件系统中的目录结构有：单级目录结构、二级目录结构、树形目录结构和非循环图目录结构，UNIX采用非循环图目录结构，即带链接的树形目录结构。

文件系统目录结构 优点 缺点

单级目录结构 简单、能实现按名存取 查找速度慢；不允许重名；不便于共享

二级目录结构           允许重名；提高了检索目录的速度 仍不利于文件共享

树形目录结构           文件的层次和隶属关系清晰，便于 只能在用户级对文件进行临时共享

实现不同级别的存取保护和文件系统

的动态装卸

非循环图目录结构 具有树形结构的优点，而且实现对 管理较复杂

文件的永久共享

**8常用的磁盘空闲区管理技术有哪几种？试简要说明各自的实现思想。**

答：常用的磁盘空闲区管理技术有：空闲空间表达法、空闲块链接法、位示图法和空闲块成组链接法。

空闲空间表法——所有连续的空闲盘块在表中占据一项，其中标出第一个空闲块号和该项中所包含的空闲块个数，以及相应的物理块号。利用该表可进行盘块的分配和文件的删除时盘块的回收

空闲块链接法——所有的空闲盘块链在一个队列中，用一个指针（空闲区头）指向第一个空闲块，而各个空闲块中都含有下一个空闲块的块号，最后一块的指针项计为NULL，表示链尾。分配和释放盘块都在链首进行

位示图法——利用一串二进制的值来反映磁盘空间的分配情况，每个盘块都对应一位。如果盘块是空闲的，对应位是0；如盘块已分出去，则对应位是1。

空闲块成组链法——把所有空闲盘块按固定数量分组，组与组之间形成链接关系，最后一组的块号（可能不满一组）通常放在内存的一个专用栈结构中。这样，对盘块的分配和释放是在栈中进行（或构成新的一组）.

**9什么是文件共享？文件链接如何实现文件共享？**

答：文件的共享是指系统允许多个用户（进程）共同使用某个或某些文件。

文件链接是给文件起别名，即将该文件的目录项登记在链接目录中。这样，访问该文件的路径就不只一条。不同的用户（进程）就可以利用各自的路径来共享同一文件。

**10什么是文件后备？数据转储方法有哪两种？按时间划分，后备分哪几种？**

答：文件的后备就是把硬盘上的文件转储到其他外部介质上。

将磁盘上的数据转储到磁带上有两种方式：物理转储和逻辑转储。物理转储是从磁盘上第0块开始，把所有的盘块按照顺序写到磁带上，当复制完最后一块时，转储结束。逻辑转储方式是从一个或多个指定的目录开始，递归地转储自某个日期以来被修改过的所有文件和目录。

通常有以下三种备份策略：完全备份、增量备份和更新备份。

完全备份也称简单备份，即每隔一定时间就对系统作一次全面的备份；增量备份使每隔一段较短的时间进行一次备份，但仅仅备份在这段时间间隔内修改过的数据；更新备份是备份从上次进行完全备份后至今更改的全部数据文件。

**11文件系统的一般格式是怎样的？其中引导块和超级块的作用各是什么？**

答：文件系统的一般由引导块、超级块、空闲空间管理、I节点、根目录、文件数据区

引导块的作用是引导操作系统。它包括一个小程序，用于读入该分区上相应操作系统的引导部分，从而把该分区中的操作系统装入内存。

超级块的作用是对整个文件系统进行控制和管理。它包含有关文件系统的全部关键参数。当计算机加电进行引导或第一次遇到该文件系统时，就把超级块中的信息读入内存。超级块中包含标识文件系统类型的幻数、文件系统中的盘块数量、修改标记及其他关键管理方面的信息。

## 第七章  设备管理

### 名词解释

**1输入井**

是指为使设备与cpu速度相匹配，系统在磁盘上设置的多个缓冲区，以实现设备与cpu之间的数据交换。输入井主要用来存放由输入设备输入的信息。

**2缓冲池**

又叫公共缓冲区，也是系统在磁盘上设置的多个缓冲区。它既可以用于输入，也可以用于输出，较好地克服了专用缓冲区的缺点。一方面提高了缓冲区的利用率，另一方面也提高了设备与cpu的并行操作程度。

**3虚拟设备**

它是利用共享设备上的一部分空间来模拟独占设备的一种I/O技术。

**4存储设备**

它们是指计算机用来存储信息的设备，如此盘（硬盘和软盘）、磁带等。

**5输入输出设备**

是计算机用来接收来自外部世界信息的设备，或者将计算机加工处理好的信息送向外部世界的设备。例如键盘、打印机、卡片输入机。

**6设备的无关性**

也称设备独立性，就是说，用户程序应与实际使用的物理设备无关，由操作系统来考虑因实际设备不同而需要使用不同的设备驱动程序等问题。

**7通道**

为使CPU摆脱繁忙的I/O事务，现代大、中型计算机都设置了专门处理I/O操作的机构，这就是通道。

**8 RAID**

称作廉价磁盘冗余阵列，即利用一台磁盘阵列控制器来统一管理和控制一组磁盘驱动器（几台到几十台），组成一个高可靠性、快速大容量的磁盘系统。采用该技术可以获取更高的可靠性和更快的数据传输速率，而不是价格上更便宜。

**简答题**

**1为什么要引入缓冲技术？设置缓冲区的原则是什么？**

答：引入缓冲区的主要目的是：⑴缓和CPU与I/O设备间速度不匹配的矛盾。⑵提高它们之间的并行性。⑶减少对CPU的中断次数，放宽CPU对中断响应时间的要求。

    设置缓冲区的原则是：如果数据到达率与离去率相差很大，则可采用单缓冲方式；如果信息的输入与输出速率相同（或者相差不大）时，则可用双缓冲区；对于阵发性的输入/输出，可以设立多个缓冲区。

**2操作系统中设备管理的功能是什么？**

答：对各种外部设备进行管理是操作系统的一个重要任务，也是其基本组成部分。

操作系统中设备管理的功能是：

① 监视设备状态；

② 进行设备分配；

③ 完成I/O操作；

④ 缓冲管理与地址转换。

**3什么是缓冲？为什么要引入缓冲？**

答：缓冲即是使用专用硬件缓冲器或在内存中划出一个区域用来暂时存放输入输出数据的器件。

引入缓冲是为了匹配外设和cpu之间的处理速度，减少中断次数和cpu的中断处理时间，同时解决dma或通道方式时的数据传输瓶颈问题。

**4 I/O设备通常可分为哪两大类？各自传输的信息单位有什么特点？**

答：I/O设备通常可分为字符设备和块设备。

字符设备通常以独占方式分配，信息的传输单位是字符或字节。块设备通常采用共享方式分配，信息的传输是以块为单位进行传输的。

**5什么是I/O控制？,I/O操作的四种控制方式是什么？**

答：I/O控制是指从用户进程的输入/输出请求开始，给用户进程分配设备和启动有关设备进行I/O操作，并在I/O操作完成之后响应中断,直至善后处理为止的整个系统控制过程 。

I/O操作的四种控制方式分别是：程序直接控制方式、中断I/O控制方式、DMA控制方式、I/O通道控制方式 。

**6 I/O控制可用那几种方式实现，各有什么优缺点？**

答：I/O控制过程可用三种方式实现：作为请求I/O操作的进程实现；作为当前进程的一部分实现；由专门的系统进程——I/O进程完成。

第一种方式请求对应I/O操作的进程能很快占据处理机，但要求系统和I/O操作的进程具有良好的实时性。第二种方式不要求系统具有高的实时性，但I/O控制过程要由当前进程负责。第三种方式增加了一个额外的进程开销，但用户不用关心I/O控制过程。

**7设备分配技术主要有哪些？常用的设备分配算法是什么？**

答：设备分配技术主要有：独占分配、共享分配和虚拟分配。

常用的设备分配算法是：先来先服务算法和优先级高的优先服务算法。

**8实现SPOOLing系统的硬件前提是什么？SPOOLing系统的主要功能是什么？**

答：实现SPOOLING系统的首先要有硬件支持：要提供大容量的磁盘，要有中断和通道装置，以便使外围设备与中央处理器能够并行工作。 它是为了满足多道程序或多进程队独占设备的共享使用而引入的，其主要功能即是：将独占设备改造为共享设备，实现虚拟设备 。

**9简述处理I/O请求的主要步骤。**

答：处理I/O请求是一个系统获取用户I/O请求转发给相应外设完成的过程 ，其具体的处理步骤如下：

①用户进程发出I/O操作；

②系统接受这个I/O请求，转去执行操作系统的核心程序；

③设备驱动程序具体完成I/O操作；

④I/O完成后,系统进行I/O中断处理，然后用户进程重新开始执行

**10设备驱动程序主要执行什么功能？**

答：设备驱动进程严格执行设备驱动程序中规定的各种功能 ，即接受用户的I/O请求 ；取出请求队列中队首的请求，将相应的设备分配给它 ；启动该设备工作，完成指定的I/O操作 ；处理来自设备的中断 。

**11 I/O软件的设计目标？它是如何划分层次的？各层的功能是什么？**

答：I/O软件的设计目标：

①与设备无关

②对文件和设备应统一命名

③层次结构

④效率高

I/O软件可分为如下4个层次：中断处理程序、设备驱动程序、与设备无关的操作系统软件和用户级软件 。各层功能为：

①中断处理程序——分析中断原因，并依据中断原因调用相应的处理程序

②设备驱动程序——它接受来自上层、与设备无关软件的抽象读写请求，并将该I/O请求排在请求队列的队尾，还要检查I/O请求的合法性；取出请求队列中对首请求，将相应设备分配给它；向该设备控制器发送命令，启动该设备工作，完成指定的I/O操作；处理来自设备的中断

③与设备无关的操作系统软件——其基本功能是执行所有驱动器共同的I/O功能和对用户级软件提供统一软件

④用户级软件——多数I/O软件都在操作系统中，用户空间中也有一小部分。通常，它们以库函数形式出现，在用户程序中可以调用它们

**12什么叫寻道？访问磁盘时间由哪几部分组成?其中哪一个是磁盘调度的主要目标？为什么？**

答:  把磁头从当前位置移到相应的磁道上或柱面上，这个操作过程叫做寻道。

访问磁盘一般要有三部分时间：寻道时间、旋转延迟时间和传输时间。

寻道时间是磁盘调度的主要目标。

传输时间是硬件设计时就固定的，寻道时间和旋转延迟时间与信息在磁盘上的位置有关。因为磁头臂是机械移动，所以对大多数磁盘来说，寻道时间远大于旋转延迟时间与传输时间之和，它是影响磁盘调度的主要因素。

**13什么是RAID？采用RAID技术的优点是什么？**

答：RAID称作廉价磁盘冗余阵列，即利用一台磁盘阵列控制起来统一管理和控制一组磁盘驱动器（几台到几十台），组成一个高可靠性、快速大容量的磁盘系统。

采用RAID技术可以获取更高的可靠性和更快的数据传输速率，而不是价格上更便宜。

**综合题**

**1假定一个硬盘有100个柱面，每个柱面有10个磁道，每个磁道有15个扇区。当进程要访问磁盘的12345扇区时，计算磁盘的三维物理扇区号。**

解：

每个柱面的扇区数为：10\*15=150  (3’)

12345/150=82余45,故12345扇区所在的柱面为82  (3’)

再将45/15，其商为3，余数为0，(3’)

故求得12345扇区所在的磁盘地址为：82柱面，3磁道，0扇区。(1’)

**2假设移动头磁盘有200个磁道（0-199号）。目前正在处理143号磁道上的请求，而刚刚处理结束的请求是125号，如果下面给出的顺序是按FIFO排成的等待服务队列顺序：86，147，91，177，94，150，102，75，130   那么，下列各种磁盘调度算法来满足这些请求所需的总磁头移动量是多少？ 假设移动头磁盘有200个磁道（0-199号）。目前正在处理143号磁道上的请求，而刚刚处理结束的请求是125号，如果下面给出的顺序是按FIFO排成的等待服务队列顺序：86，147，91，177，94，150，175，130**

**1）FCFS**

**2)SSTF**

解：

1） FCFS:125à143—86—147—91—177—94—150—102—75—130    (2’)

满足这些请求所需要的总磁头移动量为

(143-86)+(147-86)+(147-91)+(177-91)+(177-94)+(150-94)+(150-102)+(102-75)+(130-75)

=57+61+65+86+83+56+48+27+55=529   (3’)

2) SSTF:125à143—147—150—130—102—94—91—86—75—177   (2’)

满足这些请求所需的总磁头移动量为

(150-143)+(150-75)+(177-75)=7+75+102=182  (3’)

**3假设移动头磁盘有200个磁道（0-199号）。目前正在处理143号磁道上的请求，而刚刚处理结束的请求是125号，如果下面给出的顺序是按FIFO排成的等待服务队列顺序：86，147，91，177，94，150，102，75，130   那么，下列各种磁盘调度算法来满足这些请求所需的总磁头移动量是多少？**

**1)SCAN**

**2)C-LOOK**

解：

1)SCAN:125à143—147—150—177—199—130—102—94—91—86—75   (2’)

(199-143)+(199-75)=56+124=180   (3’)

2)C-LOOK:125à143—147—150—177 —75—86—91—94—102—130   (2’)

满足这些请求所需要的总磁头移动量为

177-143+177-75+130-75=33+102+55=190   (3’)

**4假设一个磁盘由200个磁道，编号从0~199。当前磁头正在143道上服务，并且刚刚完成了125道的请求。如果寻道请求队列的顺序是：86，147，91，177，94，150，102，175，130 问：为完成上述请求，下列算法各自磁头移动的总量是多少？**

1. **FCFS**
2. **SSTF**

解：

⑴FCFS磁头移动顺序:

143 à 86 à 147 à 91 à 177 à 94 à 150 à 102 à 175 à 130    (2’)

   57   61    56   86    83   56    48    73    45

磁头移动总量: 57+61+56+86+83+56+48+73+45=565    (3’)

⑵SSTF磁头移动顺序

143 à 147 à 150 à 130 à 102 à 94 à 91 à 86 à 175 à 177  (2’)

   4     3     20    28    8    3    5    89    2

磁头移动总量: 4+3+20+28+8+3+5+89+2=162     ( 3’)

**5假设一个磁盘由200个磁道，编号从0~199。当前磁头正在143道上服务，并且刚刚完成了125道的请求。如果寻道请求队列的顺序是：86，147，91，177，94，150，102，175，130 问：为完成上述请求，下列算法各自磁头移动的总量是多少？**

1. **SCAN**
2. **C-SCAN**

解：

（1）SCAN磁头移动顺序

143 à 147 à 150 à 175 à 177 à 199 à130 à 102 à 94 à 91 à 86     (2’)

   4     3     25     2     22   69    28    8    3    5

磁头移动总量: 4+3+25+2+22+69+28+8+3+5=169   (3’)

（2）C-SCAN磁头移动顺序

143 à 147 à 150 à 175 à 177 à 199 à 0 à 86 à 91 à 94 à 102 à 130    (2’)

   4     3     25    2     22    199  86  5    3    8     28

磁头移动总量: 4+3+25+2+22+199+86+5+3+8+28=385  (3’)

**6磁盘请求以10，22，20，2，40，6，38柱面的次序到达磁盘驱动器。寻道时每个柱面移动需要6ms，计算以下寻道次序和寻道时间：**

**①先到先服务        ②电梯调度算法（起始移动向上）**

**所有情况下磁臂的起始位置都是柱面20。**

解:

寻道时间=柱面（磁道）移动总量×6ms

1） 先到先服务算法的调度顺序：

20 à 10 à 22 à 20 à 2 à 40 à 6 à 38  (2’)

  10   12   2    18  38   34  32

柱面移动总量=10+12+2+18+38+34+32=146   (2’)

寻道时间=146×6ms=876ms   (1’)

2） 电梯算法的调度顺序：

20 à 22 à 38 à 40 à 10 à 6 à 2     (2’)

  2    16   2    30   4   4

柱面移动总量=2+16+2+30+4+4=58   (2’)

寻道时间=58×6ms=348ms  (1’)

**7某系统文件存储空间共有80个柱面，20磁道/柱面，6块/磁道，每块有1K字节。用位示图表示。每张位示图为64个字，其中有4个包含的是控制信息。位示图中的位若为1，表示占用；为0表示空闲。试给出分配和回收一个盘块的计算公式。**

解：每个柱面的块数为：20\*6=120（块）(1’)

计算该文件系统的磁盘块：80\*120=9600（块）(1’)

每张位示图为64个字，其中有4个字包含的是控制信息。假定每个字为16位。每张可以记录的块数为：(64-4)\*16=960（块）(1’)

总共有9600块，用位示图表示，需要占用9600位，每张位示图可记录960块，需要的位示图数位：9600/960=10（张），用0~9进行编号。(1’)

1） 分配一个盘块的计算公式 (3’)

相对块号=位图的张号\*960+字号\*16+位号

柱面号=（相对块号/每个柱面的块数）的商=（相对块号/120）的商

磁盘号=（相对块号/每个柱面的块数）的余数的商

扇区号=（（相对块号/每个柱面的块数）的余数）的余数

2） 回收一个盘块的计算公式 (3’)

先将三维地址转换为相对块号，再将相对块号转换为位图的字号和位号。

相对块号=柱面号\*120+磁盘号\*6+扇区号

字号=（相对块号/16）的商

位号=（相对块号/16）的余数

## 第八章  中断和信号机制

### 名词解释

**1 中断**

是指CPU对系统发生的某个事件做出的一种反应，CPU暂停正在执行的程序，保留现场后自动地转去执行相应的处理程序，处理完该事件后，如被中断进程的优先级最高，则返回断点继续执行被“打断”的程序。

**2中断源**

引起中断的事件或发出中断请求的来源称为中断。

**3中断请求**

中断源向CPU提出进行处理的请求。

**4中断向量**

通常包括相应中断处理程序入口地址和中断处理时处理机状态字。

**5异常**

它是指来自cpu内部的事件或程序执行中的事件引起的中断

**6程序性中断**

是指因错误地使用指令或数据而引起的中断，用于反映程序执行过程中发现的例外情况，例如，非法操作码，无效地址、运算溢出，等等。

**7断点**

发生中断时，被打断程序的暂停点称为断点。

**8中断响应**

发生中断时，cpu暂停执行当前的程序，转去处理中断。这个由硬件对中断请求做出反应的过程，称为中断响应。

**9中断屏蔽**

是指在提出中断请求之后，cpu不予响应的状态。它常常用来在处理某个中断时防止同级中断的干扰，或在处理一段不可分割的、必须连续执行的程序时防止意外事件把它打断。

**10中断禁止**

是指在可引起中断的事件发生时系统不接收该中断的信号，因而就不可能提出中断请求而导致中断。简言之，就是不让某些事件产生中断。

**11软中断**

又称信号机制，它是在软件层次上对中断机制的一种模拟，其中，信号的发送者相当于中断源，而接收者（必定是一个进程）相当于cpu。

**简答题**

**1什么是中断？什么叫中断处理？什么叫中断响应？**

答：中断是指计算机在执行期间，系统内发生任何非寻常的或非预期的急需处理事件，使得cpu暂时中断当前正在执行的程序而转去执行相应的事件处理程序，待处理完毕后又返回原来被中断处继续执行的过程。

Cpu转去执行相应的事件处理程序的过程称为中断处理。

Cpu收到中断请求后转到相应的事件处理程序称为中断响应。

**2什么叫关中断？什么叫开中断？什么叫中断屏蔽？**

答：把cpu内部的处理机状态字psw的中断允许位清除从而不允许cpu响应中断叫做关中断。

设置cpu内部的处理机状态字psw的中断允许位从而允许cpu响应中断叫做开中断。

中断屏蔽是指中断请求产生之后，系统用软件方式有选择地封锁部分中断而允许其余部分的中断仍能得到响应。

**3什么是陷阱？什么是软中断？**

答：陷阱指处理机和内存内部产生的中断，它包括程序运算引起的各种错误，如地址非法、检验错、页面失效。存取访问控制错、从用户态到核心态的切换等都是陷阱的例子。

软中断是通信进程之间用来模拟硬中断的一种信号通信方式。

**4中断响应主要做哪些工作？由谁来完成？**

答：中断响应主要做的工作是：

①中止当前程序的执行；

②保存原程序的断点信息（主要是程序计数器PC和程序状态寄存器PS的内容）；

③转到相应的处理程序

中断响应由硬件实施。

**5叙述缺页中断和一般中断的区别？**

答：缺页中断也是中断的一种，既然是中断都应保护当前运行程序的现场信息，中断完成后恢复被中断的现场。但缺页中断是当前运行进程自己产生的中断，且当前指令还未执行完，故中断处理将所需的页调入主存后，应该恢复该进程重新执行被中断的这条指令。

对一般中断，中断源与当前正在执行的进程无关，故正在执行的进程执行完当前这条指令后才响应中断。中断处理完成后，可能恢复被中断的进程，也可能调度其他进程的运行。即使恢复被中断进程的运行，恢复后执行的指令也是中断发生时的下一条指令。

**6什么是软中断？**

答：软中断是对硬中断的一种模拟，发送软中断就是向接收进程的proc结构中的相应项发送一个特定意义的信号 。软中断必须等到接收进程执行时才能生效 。

**7进程在什么时候处理它接收到的软中断信号？进程接收到软中断信号后放在什么地方？**

答：进程在再次被调度执行时先检查是否收到软中断，若进程接收到了软中断信号则优先处理软中断 。进程把接收到软中断信号存放在proc结构的相应项中 。

**8中断处理的主要步骤是什么？**

答：中断处理的一般步骤是:

保存被中断程序的现场，

分析中断原因，

转入相应处理程序进行处理，

恢复被中断程序现场（即中断返回）。

**9什么叫系统调用？执行用户程序中的系统调用时，相应进程的状态会发生什么变化？**

答：系统调用是用户在程序中能以“函数调用”形式调用的、由操作系统提供的子功能的集合。每一个子功能称作一条系统调用命令。它是操作系统对外的接口，是用户级程序取得操作系统服务的唯一途径。

执行到用户程序中的系统调用时，相应进程的状态从用户态变为核心态。

**10在用户程序执行过程中，CPU接到盘I/O中断。对此，系统（硬件和软件）要进行相应处理，试列出其主要处理过程。**

答：硬件主要处理过程是：cpu中止当前程序的正常执行；保存原程序的程序计算器pc和程序状态寄存器ps的内容：取出盘I/O中断向量，转到相应的处理程序。

软件主要处理过程是：保存被中断程序的现场（如通用寄存器的内容）：分析中断原因，由中断向量得到盘I/O中断的处理程序地址；运行盘I/O中断处理程序，判断I/O工作是否完成，如正常完成，则作I/O结束处理；执行完中断处理程序，核心恢复前面保存的现场，进程回到用户态。