**第1章   操作系统概论**

**1 早期操作系统设计的主要目标是什么?**

方便性：方便用户使用计算机。用户通过操作系统来使用计算机。

有效性：使计算机系统能高效可靠地运转，提高系统资源的利用率。

还要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？**

进程----进程表  存储器----存储表  I/O设备----I/O设备表  文件----文件表

**3 为什么要引入多道程序系统?它有什么特点？**

提高CPU的利用率，充分发挥系统设备的并行性。这包括程序之间、CPU与设备之间、设备与设备之间的并行操作。 指在主存同时存放若干道程序，使它们在系统中交叉运行，共享系统中的各种资源。当一道程序暂停执行时，CPU立即转去执行另一道程序。

在单处理机系统中。宏观上，多道程序并行运行；微观上，在任何特定时刻，只有一道程序在处理机上运行，即各程序交叉地在CPU上运行。

**4 叙述操作系统的基本功能。**

(1)处理机管理：进程管理。处理机如何调度的问题：FCFS、优先级、时间片轮转？

(2)存储器管理：主存管理。存储分配、存储保护、主存扩充。

(3)设备管理：涉及对系统中各种输入、输出设备的管理和控制。分配设备，控制设备传输数据。

(4)文件管理：将程序、数据、操作系统软件等组织成文件，存在磁盘或磁带上，方便用户访问。

**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?**

**批处理**：优点: 系统吞吐量大，资源利用率高。适合计算量大、自动化程度高的成熟作业。

缺点: 用户与作业无法交互，作业平均周转时间较长。

适用于比较成熟的大作业

**分时系统**：同时性：若干用户同时使用一台计算机。

独立性：每个用户占有一台终端，独立操作，感觉不到别的用户存在。

交互性：用户可通过终端与系统进行人机对话。

及时性：用户的请求能在较短时间内得到响应。

适用于小作业

**实时系统：**（1）实时性。其响应时间由被控制对象所能承受的延迟来确定。

（2）可靠性。要具有容错能力，可采用双工机制：一台主机；一台后备机。

（3）确定性。是指系统按照固定的、预先确定的时间执行指定的操作。其可确定性取决于系统响应中断的速度和处理能力。

适用于实时过程控制，实时信息处理

**6 操作系统的特性?**

（1）并发性：并发是指系统中存在着若干个逻辑上相互独立的程序，它们都已被启动执行，都还没有执行完，并竞争系统资源。

(2)共享性：是指系统中的资源可供内存中多个并发执行的进程共同使用。如打印机、磁带机、磁盘等。 支持系统并发性的物质基础是资源共享

(3)虚拟性：把共享资源的一个物理实体变为若干个逻辑上的对应物。如，CPU的分时共享；虚拟存储器技术。

(4)异步性（随机性）：有限的资源共享使并发进程之间产生相互制约关系。各个进程何时执行、何时暂停、以怎样的速度向前推进、什么时候完成等都是不可预知的。

**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？**

资源利用率：指在给定时间内，系统中某一资源（如CPU、存储器、外部设备等）实际使用时间所占比率。

吞吐量(Throughput)：指单位时间内系统所处理的信息量。它通常是用每小时或每天所处理的作业个数来度量。

周转时间：指从作业进入系统到作业退出系统所用的时间。而平均周转时间是指系统运行的几个作业周转时间的平均值。

**8  什么是嵌入式系统？**

以实际应用为中心、以计算机技术为基础、软硬件可裁剪的专用计算机系统。

软件要求固化存储。通常是一个多任务可抢占式的实时操作系统，只有满足实际需要的有限功能，如任务调度、同步与通信、主存管理、时钟管理等。

**9  什么是对称多处理？它有什么好处？**

多处理机采用紧耦合方式进行连接，共享主存

a) 非对称多处理(ASMP)：主处理机运行操作系统，其他处理机运行用户作业，主处理机为其他处理机分配和调度任务，主从模式。

b) 对称多处理(SMP)：操作系统和用户程序可安排在任何一个处理机上运行，各处理机共享主存和各种I/O设备。

**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？**

CPU的运行状态分为核心态（管态）和用户态（目态）。通过中断和异常，CPU能从用户程序的运行转入操作系统内核程序的运行。在核心态下， 允许执行处理机的全部指令集，访问所有的寄存器和存储区；在用户态下，只允许执行处理机的非特权指令，访问指定的寄存器和存储区。用户态到核心态的转换由硬件完成；管态到目态的转换由操作系统程序执行后完成。

**11  什么是系统调用，特权指令？特权指令执行时，CPU处于哪种工作状态？**

系统调用就是操作系统内核提供的一些子程序。操作系统内核向用户提供了一组系统调用接口。用户通过系统调用接口，向操作系统提出资源请求或获得系统服务。特权指令是指关系系统全局的指令。如存取和操作CPU状态。启动各种外部设备，设置时钟时间，关中断，清主存。只允许操作系统使用，不允许用户使用CPU指令集分为特权指令和非特权指令。特权指令执行时，CPU处于和心态。

**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?**

操作接口：命令语言或窗口界面是用户使用计算机系统的主要接口。

编程接口：系统调用是用户与操作系统之间的编程接口。

**第2-3章  进程管理**

**1 程序顺序执行的特点**

程序的顺序（串行）执行：计算机每次只运行一道程序。如单道批处理系统。

封闭性：程序在运行时独占全机资源，因此，这些资源的状态只能由该程序决定和改变，不受外界因素影响。

可再现性：只要初始条件相同，无论程序连续运行，还是断断续续地运行，程序的执行结果不变。

优点： 由于顺序程序的封闭性和可再现性，为程序员调试程序带来了很大方便。

缺点： 由于资源的独占性，使得系统资源利用率非常低。

**2 何谓进程，进程由哪些部分组成？试述进程的四大特性及进程和程序的区别。**

操作系统用“进程” 来描述系统中各并发活动进程（process）又叫做任务（task）进程是程序的一次执行过程进程是程序在一个数据集合上顺序执行时发生的活动

动态性。进程是程序的一次执行过程，是临时的，有生命期的。

独立性。进程是系统进行资源分配和调度的一个独立单位。

并发性。多个进程可在处理机上交替执行。

结构性。系统为每个进程建立一个进程控制块。

进程是动态的，程序是静态的。程序是有序代码的集合，进程是程序的执行，没有程序就没有进程。通常，进程不可以在计算机之间迁移，而程序可以复制。

进程是暂时的，程序是永久的。

进程包括程序、数据和进程控制块。

通过多次执行，一个程序可对应多个进程；通过调用关系，一个进程可包括多个程序。进程可创建其他进程，而程序不能形成新的程序。

**3 进程控制块的作用是什么?它主要包括哪几部分内容?**

也叫进程描述符，是为了描述进程的运行变化情况，操作系统为每个进程定义了一个数据结构，它是进程存在的唯一标识。它包含了进程的描述信息和管理控制信息。具体包括：

1. 进程标识数：用于唯一地标识一个进程，通常是一个整数。

  外部标识符，由用户使用。如：send进程、print进程等。

2. 进程的状态、调度、存储器管理信息：是调度进程所必需的信息，包括进程状态、优先级、程序在主存地址、在外存的地址等。

3. 进程使用的资源信息：分配给进程的I/O设备、正在打开的文件等。

4. CPU现场保护区：保存进程运行的现场信息。包括：程序计数器(PC)、程序状态字、通用寄存器、堆栈指针等。

5. 记帐信息：包括使用CPU时间量、帐号等。

6. 进程之间的家族关系：类UNIX系统，进程之间存在着家族关系，父/子进程。Windows 进程之间不具有父子关系。

7. 进程的链接指针：链接相同状态的进程。

**4 进程的基本状态，试举出使进程状态发生变化的事件并描绘它的状态转换图。**

（1）运行态(running)：进程正在CPU上运行。单CPU系统一次只有一个运行进程；多CPU系统可能有多个运行进程。

(2)阻塞态(blocked)：又称等待态。当进程因等待某个条件发生而不能运行时所处的状态。等待I/O完成，等待一个消息

(3)就绪态(ready)：已获得除CPU之外的全部资源，只要再获得CPU，就可执行。

(4) 创建态：刚刚建立，未进就绪队列。

(5) 终止态：已正常结束或故障中断，但尚未撤消。暂留在系统中，方便其它进程去收集该进程的有关信息。

**5 什么是原语?什么是进程控制？**

原语是由若干多机器指令构成的完成某种特定功能的一段程序，具有不可分割性.即原语的执行必须是连续的，在执行过程中不允许被中断

进程控制：是指系统使用一些具有特定功能的程序段来创建、撤消进程，以及完成进程各状态之间的转换。进程控制是由操作系统内核实现的。是属于原语一级的操作，不能被中断。

**6 进程调度的功能、方式、时机、算法。作业调度，交换调度。作业的周转时间和作业的带权周转时间?**

**功能：**(1) 记录系统中各进程的执行状况。管理进程控制块，将进程的状态变化及资源需求情况及时地记录到PCB中。

(2) 选择就绪进程占有CPU

(3) 进行进程上下文的切换。将正在执行进程的上下文保存在该进程的PCB中，将刚选中进程的运行现场恢复起来，以便执行。

**方式：**① 非抢先方式(非剥夺方式)  某一进程占用CPU,直到运行完或不能运行为止，其间不被剥夺。用在批处理系统。主要优点：简单、系统开销小。

② 抢先方式(剥夺方式) 允许调度程序基于某种策略（优先级、时间片等）剥夺现行进程的CPU给其它进程。用在分时系统、实时系统。

**时机：**（1）现行进程完成或错误终止；

（2）提出I/O请求，等待I/O完成时；

（3）在分时系统，按照时间片轮转，分给进程的时间片用完时；

（4）优先级调度，有更高优先级进程就绪；

（5）进程执行了某种原语操作，如阻塞原语和唤醒原语，都可能引起进程调度。

**算法：**进程调度所采用的算法是与整个系统的设计目标相一致的。

**批处理系统**：增加系统吞吐量和提高系统资源的利用率;

**实时系统**：保证对随机发生的外部事件做出实时响应。

**作业调度（1-3）：**（1）先来先服务调度法(FCFS)：简单，节省机器时间。缺点：容易被大作业垄断，使得平均周转时间延长。

（2）最短作业优先调度法(SJF)：选取运行时间最短的作业运行。对短作业有利，作业的平均周转时间最佳。

（3）响应比高者优先调度法(HRN)

 Rp =(作业等待时间+作业估计运行时间)/作业估计运行时间 =1+作业等待时间/作业估计运行时间

交换调度（4-6）：（4）优先级调度法

将CPU分配给就绪队列中优先级最高的进程。包括：静态优先级，动态优先级。

（5）轮转法(Round Robin)

（6）多级反馈队列轮转法

**分时系统**：保证每个分时用户能容忍的响应时间。

时钟驱动法：各任务的调度安排在系统运行前就确定了，调度程序按照这个安排依次调度每一个任务执行。

加权轮转法：进程的权就是分配给它的一小部分处理机时间。轮转时，不同的进程可以获得不同的处理机时间。广泛用在高速开关网的实时控制中。

**7 线程的定义，线程与进程的比较。系统对线程的支持（用户及线程、核心级线程、两级组合）。**

线程(thread)：是进程内的一个可执行实体，是处理机调度的基本单位

比较：

**（1）拥有的资源**

进程拥有一个独立的地址空间，用来存放若干代码段和数据段。若干打开文件，以及至少一个线程。

一个进程内的多线程共享该进程的所有资源，线程自己拥有很少资源。

**（2）调度**

进程调度需进行进程上下文的切换，开销大。

同一进程内的线程切换，仅把线程拥有的一小部分资源变换了即可，效率高。同一进程内的线程切换比进程切换快得多。但是，在由一个进程的线程向另一个进程的线程切换时，会引起进程上下文的切换。

**（3）并发性**

   引入线程后，使得系统的并发执行程度更高。 进程之间、进程内的多线程之间可并发执行。

**（4）安全性**

   同一进程的多线程共享进程的所有资源，一个线程可以改变另一个线程的数据，而多进程实现则不会产生此问题。共享方便。

**8 并发执行的进程在系统中通常表现为几种关系?各是在什么情况下发生的?**

（1）对资源的共享引起的互斥关系

 进程之间本来是相互独立的，但由于共享资源而产生了关系。间接制约关系，互斥关系。

（2）协作完成同一个任务引起的同步关系

一组协作进程要在某些同步点上相互等待发信息后才能继续运行。直接制约关系，同步关系。

（3）进程之间的前序关系

由于进程之间的互斥同步关系，使得进程之间具有了前序关系，这些关系决定了各个进程创建和终止的时间。

**8 什么叫临界资源?什么叫临界区?对临界区的使用应符合的四个准则（互斥使用、让权等待、有空让进、有限等待）。**

临界资源：就是一次仅允许一个进程使用的系统中共享资源。

临界区(critical section)：就是并发进程访问临界资源的那段必须互斥执行的程序。

四条原则：互斥使用、让权等待、有空让进、有限等待

**10 解决进程之间互斥的办法：**

**开、关中断，加锁、开锁（又叫测试与设置，通常由一条机器指令完成），信号量与P、V操作。**

(1) 关中断   最简单的方法。在进程刚进入临界区后，立即禁止所有中断；在进程要离开之前再打开中断。因为CPU只有在发生时钟中断或其它中断时才会进行进程切换。

(2) 使用测试和设置硬件指令锁位变量W ：为每个临界资源设置一个，以指示其当前状态。W=0，表示资源空闲可用；W=1，表示资源已被占用。（忙等）

信号量和PV操作：两个或多个进程通过简单的信号进行合作，一个进程被迫在某一位置停止，直到它接收到一个特定的信号。为了发信号，需要使用一个称作信号量的特殊变量。

显然，P、V操作的引入，克服了加锁操作的忙等待现象，提高了系统的效率。操作系统正是利用信号量的状态来对进程和资源进行管理的。

**11 若信号量S表示某一类资源，则对S执行P、V操作的直观含意是什么? 当进程对信号量S执行P、V操作时，S的值发生变化，当S>0、S=0、和S<0时，其物理意义是什么?**

**12 在用P/V操作实现进程通信时，应根据什么原则对信号量赋初值？**

**13 经典的IPC问题。**

**14 进程高级通信有哪些实现机制？**

高级通信：是指进程采用系统提供的多种通信方式来实现通信。如消息缓冲、信箱、管道、共享主存区等。

发送进程和接收进程的消息通信方式：

非阻塞发送，阻塞接收

非阻塞发送，非阻塞接收

阻塞发送，阻塞接收

**15  死锁产生的必要条件及解决死锁的方法**

一组进程是死锁的，是指这一组中的每个进程都正在等待该组中的其他进程所占用的资源时，可能引起的一种错误现象。

互斥条件。独占性的资源。

保持和等待条件。进程因请求资源而阻塞时，对已经获得的资源保持不放。

不剥夺条件。已分配给进程的资源不能被剥夺，只能由进程自己释放。

循环等待条件。存在一个进程循环链，链中每个进程都在等待链中的下一个进程所占用的资源。

系统资源配置不足，引起进程竞争资源。

并发进程请求资源的随机性，包括所请求资源的类别和数量。

各并发进程在系统中异步向前推进，造成进程推进顺序的不合理性。

产生死锁的根本原因：是对独占资源的共享，并发执行进程的同步关系不当。

死锁的解决办法：

① 鸵鸟算法。忽略死锁。

② 死锁的预防。通过破坏产生死锁的四个必要条件中的一个或几个，来防止发生死锁。

③ 死锁的避免。是在资源的动态分配过程中，用某种方法去防止系统进入不安全状态，从而避免发生死锁。

④ 死锁的检测和恢复。允许死锁发生，通过设置检测机构，及时检测出死锁的发生，然后采取适当措施清除死锁。

**16 理解银行家算法的实质。能够利用银行家算法避免死锁。**

**第4章  存储器管理**

**1 存储器管理的功能。名字空间、地址空间、存储空间、逻辑地址、物理地址。**

（1）存储器分配：解决多道程序或多进程共享主存的问题

（2）地址转换或重定位：研究各种地址变换方法及相应的地址变换机构。

（3）存储器保护：防止故障程序破坏OS和其它信息

（4）存储器扩充：采用多级存储技术实现虚拟存储器及所用的各种管理算法。

（5）存储器共享：并发执行的进程如何共享主存中的程序和数据。

**2 什么是地址重定位?分为哪两种?各是依据什么和什么时候实现的?试比较它们的优缺点。**

（1）把程序地址空间的逻辑地址转换为存储空间的物理地址。静态重定位和动态重定位

静态：在进程执行前，由装入程序把用户程序中的指令和数据的逻辑地址全部转换成存储空间的物理地址。

静态重定位特点：

1）无硬件变换机构；2）为每个程序分配一个连续的存储区；3）在程序执行期间不能移动，主存利用率低；4）难以做到程序和数据的共享；5）用于单道批处理系统。

动态：装入程序把程序和数据原样装入到已分配的存储区中。程序运行时，把该存储区的起始地址送入重定位寄存器。需硬件地址转换机构。多道批处理系统、分时系统

优点：

1）主存利用充分。可移动用户程序。移动后，只需修改重定位寄存器。2）程序不必占有连续的存储空间。 3）便于多用户共享存储器中的同一程序和数据

**3   内存划分为两大部分：用户空间和操作系统空间。存储器管理是针对用户空间进行管理的。**

通常存储器划分为两部分：一部分是操作系统占用区，另一部分是用户进程占用区（或用户区）。 存储器管理是指对用户区的管理。

**4 存储保护的目的是什么?对各种存储管理方案实现存储保护时，硬件和软件各需做什么工作?**

防止地址越界：进程运行时产生的所有存储器访问地址都要进行检查，确保只访问为该进程分配的存储区域。

正确地进行存取：对所访问的存储空间的操作方式（读、写、执行）进行检查，以防止由于误操作，使其数据的完整性受到破坏。

**5 试述可变式分区管理空闲区的方法及存储器的保护方式。覆盖与交换有什么特点？**

根据作业的大小动态地划分分区，使分区的大小正好等于作业大小。各分区的大小是不定的；内存中分区的数目也是不定的。

**（**1）首次适应(first fit)法：

要求空闲区表或空闲区链中的空闲区按地址从小到大排列。分配内存时，从起始地址最小的空闲区开始扫描，直到找到一个能满足其大小要求的空闲区为止。分一块给请求者，余下部分仍留在其中。

（2）最佳适应(best fit)法：

存储分配程序要扫描所有空闲区，直到找到能满足进程需求且为最小的空闲区为止。

缺点：

因为要查找所有的分区，所以比首次适应算法效率低。可能把主存划分得更小，出现很多无用的碎片。改进：从小到大对空闲区排序。

（3）最坏适应(worst fit)法：

要扫描所有的空闲区，直到找到满足进程要求且为最大的空闲区为止。一分为二，一部分分给进程，另一部分仍留在链表中。

目的：使剩下的空闲区可用。

缺点：要扫描所有的空闲区；大空闲区的不断分割，可能满足不了大进程的要求。

改进：从大到小对空闲区排序，以提高查找速度。

覆盖：是指同一主存区可以被不同的程序段重复使用。通常一个进程由若干个功能上相互独立的程序段组成，进程在一次运行时，也只用到其中的几段。让那些不会同时执行的程序段共用同一个主存区。特点：打破了必须将一个进程的全部信息装入主存后才能运行的限制。在逻辑上扩充了主存。小主存可运行大进程

交换：系统根据需要把主存中暂时不运行的进程中的部分或全部信息移到外存，而把外存中的进程移到主存，并使其投入运行。特点：打破了一个程序一旦进入主存，便一直运行到结束的限制。

**6 页表的作用是什么?简述页式管理的地址变换过程。能利用页表实现逻辑地址转换成物理地址。管理内存的数据结构有哪些？**

页表：系统为每个进程建立一张页面映像表，记录逻辑页与主存块的映射关系。

管理内存的数据结构有两种：存储分块表，位示图；

**7  什么是页式存储器的内零头?与页的大小有什么关系?可变式分区管理产生什么样的零头？**

**8   段式存储器管理与页式管理的主要区别是什么?**

(1)段是由用户划分的；页是为了方便管理由硬件划分的，对用户是透明的。

(2)页的大小固定；段的大小不固定。

(3)段式用二维地址空间；页式用一维地址空间。

(4)段允许动态扩充，便于存储保护和信息共享。

(5)段可能产生主存碎片；页消除了碎片。

(6)段式管理便于实现动态链接，页式管理只能进行静态链接。

(7)段与页一样，实现地址变换开销大，表格多。

**9 什么是虚拟存储器。虚拟存储器的容量能大于主存容量加辅存容量之和吗?**

虚拟存储器：是系统为了满足应用对存储器容量的巨大需求而构造的一个非常大的地址空间。其容量由计算机的地址结构确定。系统的指令地址部分能覆盖的地址域大于实际主存的容量。

**10 实现请求页式管理，需要对页表进行修改，一般要增加状态位、修改位。试说明它们的作用。**

**11 产生缺页中断时，系统应做哪些工作?**

1）根据当前指令的逻辑地址查页表的状态位。2）状态位为0，缺页中断。3）操作系统处理缺页中断，寻找一个空闲的页框。4）若有空闲页，则把从磁盘读入信息装入该页框。5）若无空闲页，则按某种算法选择一个已在内存的页面，暂时调出内存。若修改过还要写磁盘。调入需要的页。之后要修改相6）应的页表和内存分配表。7）恢复现场，重新执行被中断的指令。

**12 会利用FIFO、LRU、OPT以及时钟页面置换算法描述页面置换过程，计算产生的缺页率。Belady异常。**

**13 什么是程序的局部性原理？什么叫系统抖动?工作集模型如何防止系统抖动?**

时间局部性：程序中往往含有许多循环，在一段时间内会重复执行该部分。

空间局部性：程序中含有许多分支，在一次执行中，只有满足条件的代码运行，不满足条件的代码不运行。即使顺序执行程序，程序的地址域在短时间内变化不大。在进程运行过程中，用到哪部分程序或数据再由系统自动装入。

**14  多级页表的概念，多级页表中页表建立的时机。写时复制技术的概念。**

大页表：(地址空间4GB)/(页4KB)=1M个页表项。若每个页表项占4B，则最大的页表为4MB。

多级页表结构：页表在内存不必连续存放。

页表的建立不再是在进程装入主存时，而是推迟到要访问页时，才为包含该页的页表分配空间和建立页表页。

**第5章   文件系统**

**1 什么是文件和文件系统? 文件系统的主要功能。UNIX系统如何对文件进行分类？它有什么好处？**

文件是存储在外部存储器上的具有符号名的相关信息的集合。

文件系统是OS中管理文件的软件机构。包括管理文件所需的数据结构、相应的管理软件和被管理的文件。

主要功能：管理文件存储器。1）记录空间使用情况，分配空间，调整或回收空间。2）实现按名存取。利用目录结构快速定位文件。3）应具有灵活多样的文件结构和存取方法，便于用户存储和加工处理信息。4）提供一套使用方便、简单的操作命令。5）保证文件信息的安全性。6）便于文件的共享。

UNIX分为普通文件，目录文件，特殊文件

**2 文件目录的作用是什么?文件目录项通常包含哪些内容? 文件控制块。**

主要作用是使用户实现按名存取文件。主要记录文件的名字机器存放物理地址的一张映射表，表中包括了许多文件控制块（FCB）

**3 文件的逻辑结构有几种形式？文件的存取方法？**

无结构的字节流式文件。由无结构的先后到达的相关字节组成，其文件长度就是所包含的字节个数。

有结构的记录式文件。分为定长记录式文件和变长记录式文件。

（1）顺序存取：按照文件信息的逻辑顺序依次存取。是在前一次存取的基础上进行的。

（2）直接存取（随机存取）基于文件的磁盘模型，磁盘允许对任意文件块进行随机读和写。对记录式文件而言。根据记录的编号来直接存取文件中的任意一个记录。对字节流文件而言。根据系统调用命令把读/写指针调整到欲读/写位置上，然后读/写指定字节数的信息

**4 文件的物理结构有几种组织形式?对于不同的组织形式，文件系统是如何进行管理的?**

1）连续文件（顺序文件）文件内容连续存放。优点简单。支持顺序存取和随机存取。存取速度快。只要访问一次文件的管理信息，就可方便地存取到任一记录。缺点：不灵活，容易产生碎片。

2）链接文件 不要求文件内容连续存放。把文件所占用的物理块用链接指针链接起来。优点：可以解决外存的碎片问题，提高了外存空间的利用率；允许文件动态增长。 缺点：只能按文件的指针链顺序存取，查找效率较低。

3）索引文件 为每个文件建立一张索引表。用索引表记录文件内容的存放地址，即记录文件的逻辑块号和对应的物理块号之间的关系。优点：文件可动态修改；随机、顺序存取。缺点：索引表的使用增加了存储空间的开销；降低了文件的存取速度。

 4）索引顺序文件

**5 DOS文件卷的结构，DOS系统的文件物理组织是什么？**

MS-DOS就是使用文件分配表（FAT）来分配和管理磁盘空间的。DOS系统的文件采用链接结构。（索引顺序文件）

**6 为什么要进行记录的组块和分解？**

对于记录式文件，记录大小是由文件的性质决定的，而存储介质上的块的划分与存储介质的特性有关，块的大小通常不是固定不变的。因此，实际一个逻辑记录与物理块的大小是不相等的。当用户文件的逻辑记录远小于物理块大小时，一个逻辑记录存放在一个物理块总，将会造成极大的浪费。为此，可把多个记录存放在一个物理块中，也允许一个逻辑记录跨块存放。这样可以提高存储器的利用率。

**7 文件存储空间的管理方法有几种?它们各是如何实现文件存储空间的分配和回收的?**

空白文件目录 （是一种最简单的方法）系统为所有这些空白文件建立一张表。每个空白文件占用一个表目。

适合于文件的静态分配（连续文件的分配）

 空闲块链表 （1）空闲块链 把所有空闲块连接成一个链表。优点：简单。适合文件动态分配。缺点：工作效率低。分配和回收多个盘块时要多次访问磁盘才能完成。（2）空闲块成组链表利用盘空闲块来管理盘上的空闲块，每个磁盘块记录尽可能多的空闲块而成一组。各组之间也用链指针链接在一起。便于空闲块的分配与回收。适合连续文件、链接文件和索引文件的存储分配。

位映像表(bit map)或位示图是适合文件静态分配和动态分配的最简单方法每一个二进制位对应一个物理盘块。为1时表示块已分配，为0时空闲。

**8 建立多级目录有哪些好处?文件的重名和共享问题是如何得到解决的?**

**9 文件系统中，常用的文件操作命令有哪些？打开和关闭文件命令的目的是什么?它们的具体功能是什么?**

**创建(Create)文件**  主要功能：在指定设备上为指定路径名的文件建立一个目录项，并设置文件的有关属性。

**删除(Delete)文件**  主要功能：根据文件的路径名找到指定的目录项，回收其占用的各个物理块，再将该目录项置为空。

**打开(Open)文件**    根据文件路径名找到目录项，将文件的目录项复制到主存一个专门区域，返回文件在该区域的索引。建立进程与文件的联系。  目的：避免多次重复地检索文件目录。  系统维护了一个系统当前打开文件表。当读/写文件时，通过这个表的索引找到文件的主存目录项。不需要重复地对磁盘进行检索。

**关闭(Close)文件**  释放文件在主存专门区域中的目录项，切断用户与文件的联系。若该目录项被修改过，则复制到磁盘。若文件作过某些修改，应将其写回辅存。

**读(Read)文件**     命令中必须指出要读的数据个数，以及存放数据的主存地址。根据文件所在设备、文件类型的不同，系统设置不同的读命令。

**写(Write)文件**   命令中必须指出要写的数据个数，以及存放数据的主存地址，将主存中的数据写到指定的文件中。

**追加(Append)文件** 限制了写文件的形式，将数据追加到文件尾。

随机存取(Seek)文件 重新定位文件的读/写位置指针。

**得到文件属性(Get Attributes)**   进程在执行时常常需要了解文件的属性。在UNIX系统中，一个软件开发项目通常由多个源文件组成，make程序用来管理这些软件开发项目。当make被调用时，它检查所有源文件和目标文件的修改时间，并且编排出需要重新编译的文件数。

**设置文件属性(Set Attributes)**  修改文件的一些属性，以适应用户的要求。

**重命名(Rename)文件**  重新命名一个已经存在的文件。

**11 存取控制表ACL的概念。**

为存取控制矩阵中的每一列建立一张存取控制表(ACL)，用一有序对(域, 权集)表示。

**12 理解存储器映射文件。**

将文件映射到进程地址空间的一个区域，返回虚拟地址，仅当需要对文件存取时，才传输实际的数据。

**第6章   设备管理**

**1 I/O设备通常大致可分为哪两大类？各自传输的信息单位有什么特点？**

字符设备：人机交互设备。是以字符为单位发送和接收数据的，通信速度比较慢。键盘和显示器、鼠标、扫描仪、打印机等。

块设备：外部存储器。以块为单位传输数据。常见块尺寸：512B～32KB。如磁盘、磁带、光盘等。

**2 常用的四种数据传输方式。**

程序查询方式（polling）  中断方式  直接存储器访问(DMA)方式   通道控制方式

**3 根据设备的使用方式，设备被分为几种类型？何为虚拟设备？它是通过什么技术实现的？**

虚拟设备，共享设备，独占设备。

虚拟设备：是指设备本身是独占设备，而经过虚拟技术处理，可以把它改造成共享设备，供多个进程同时使用。常用可共享的高速设备来模拟独占的慢速设备。能有效提高独占型设备的利用率。 Spooling技术是实现虚拟设备的具体技术。它利用可共享磁盘的一部分空间，模拟独占的输入/输出设备。以空间换时间

**4 按照设备管理的层次结构，I/O软件划分为几层？各层主要实现哪些功能？**

中断处理程序  进程在启动一个I/O操作后阻塞起来，I/O操作完成，控制器产生一个中断。CPU响应中断，执行中断处理程序。检查设备状态。若正常完成，就唤醒等待的进程。然后检查是否还有待处理的I/O请求，若有就启动。若传输出错，再发启动命令重新传输；或向上层报告“设备错误”的信息。中断返回被中断的进程，或转进程调度。

设备驱动程序  每个设备驱动程序处理一种类型设备。由一些与设备密切相关的代码组成。提供一些与文件类似的API：open, close, read, write, control等。是OS中唯一知道设备控制器的配置情况，如设置有多少个寄存器以及这些寄存器作用。 通常包含三部分功能：①设备初始化。②启动设备传输数据的例程。③中断处理例程。

独立于设备的软件  实现所有设备都需要的功能，且向用户提供一个统一的接口。

用户空间的I/O软件

**5 何为设备的独立性？**

（设备独立性是指用户及用户程序不受系统配置的设备类型和具体设备的台号的影响。用户只是使用逻辑设备，具体的映射由操作系统完成。）

**6   什么是SPOOLING技术？以输出为例，说明它的实现原理。**

是实现虚拟设备的具体技术。它利用可共享磁盘的一部分空间，模拟独占的输入/输出设备。以空间换时间。

Spooling实际是一种缓冲技术。进程要打印时，系统并不为它分配打印机，而是把待打印的数据缓冲到一个独立的磁盘文件上，形成待打印文件队列。之后，Spooling系统一次一个地将打印队列上的文件送打印机打印。这种技术又叫缓输出技术。

**7   一个特定磁盘上的信息如何进行编址？**

盘面号、磁道号 和扇区号（或柱面号、磁头号和扇区号）。

**8  要将磁盘上一个块的信息传输到主存需要系统花费哪些时间？**

（寻道时间、旋转延迟时间和读/写传输时间）

**9 常用的磁盘调度算法：**

先来先服务、最短寻道时间优先、扫描法。（SCAN, C\_SCAN, LOOK, C\_LOOK）