

# 自考《操作系统》串讲资料

## 第一部分 课程概述

### 一、使用教材

《操作系统》，全国高等教育自学考试指导委员会组编，谭耀铭主编，中国人民大学出版社 2007 年版。

### 二、课程简介

随着计算机技术的迅速发展，计算机的硬、软件资源越来越丰富，用户也要求能更方便、更灵活地使用计算机系统。为了增强计算机系统的处理能力以及方便用户有效地使用计算机系统，操作系统已成为现代计算机系统中不可缺少的重要组成部分。因此，操作系统课程也就成为高等学校计算机专业的重要专业基础课程。

本课程从操作系统实现资源管理的观点出发，阐述如何对计算机系统内的硬、软件资源进行管理，使计算机系统协调一致地、有效地为用户服务，充分发挥资源的使用效率，提高计算机系统的可靠性和服务质量。

### 三、课程学习与考试要求

在自学过程中、考生应掌握操作系统对各种资源的管理方法和操作系统各部分之间的联系，这样才能真正掌握操作系统的工作原理以及了解操作系统在整个计算机系统中的作用。

为了能确切地学好本课程，要求考生具备：高级语言程序设计、数据结构、计算机系统结构等课程的知识。

在学任何一门课程时，要能做到真正学有所得，就必须：认真阅读教材，边读边做笔记；及时做习题，巩固所学内容；做好阶段总结，正确理解课程内容。

但每一门课程都有自身的特殊性，对于具体课程来说，应按照课程的特点具体探讨如何进行学习。要学好操作系统课程，关键在于弄清操作系统要做什么，怎么去做和为什么要这样去做。所以，如能注意如下几点，将会对操作系统课程的自学起到促进作用。

(1) 在开始阅读某一章教材之前，先翻阅大纲中有关这一章的考核知识点、考核要求、自学要求、重点，以便在阅读教材时做到心中有数，有的放矢。

(2) 阅读教材时要逐段细读，逐句推敲，集中精力，吃透每一个知识点，对基本概念必须深刻理解，对基本原理必须彻底弄清，对设计技巧要能灵活运用。

(3) 根据操作系统在计算机系统中的应用，你在学习操作系统课程时应围绕如下四个中心问题：

- ①操作系统怎样管理计算机系统中的各种资源，以及保证资源的使用效率。
- ②操作系统怎样为用户提高良好的运行环境，以方便用户使用计算机系统。
- ③操作系统怎样按照用户的要求来控制程序的执行，以保证用户得到满意的服务。
- ④操作系统怎样合理分配和管理共享资源，以确保计算机系统的安全可靠。

你在学习过程中根据这四个方面边学边总结，那么就可对操作系统有一个比较正确和全面的理解。

(4) 对于自学者来说阅读一遍书是不够的，有时阅读三遍也没完全明白，这不足为奇，也不必丧失信心。想想在校学生的学习过程，他们在课前预习，课堂听老师讲解，课后复习，在做习题等。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园....

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id, 可以直接进入俱乐部

所以,要真正学习一门课反复阅读是正常现象。

(5) 做习题是理解、消化和巩固所学知识,培养分析问题能力的重要环节,在做习题前应先认真仔细阅读教材,切忌根据习题选择教材内容,否则本末倒置,欲速则不达。

#### 四、前导课程和后继课程

操作系统是管理计算机系统资源和控制程序执行的一种系统软件。它直接扩充裸机(不配有任何软件的计算机)的功能,为程序的执行提供良好环境。所以,在学习操作系统之前应该先学习计算机组成原理、数据结构、高级语言程序设计、汇编语言程序设计等课程。在这些先修课的基础上再学习本课程,符合循序渐进的规律,不仅容易理解课程内容,而且能正确地把操作系统的各部分程序有机地联系起来。

#### 四、试题题型及分值

1. 根据最新大纲的要求操作系统试卷通常采用的题型有:单项选择题、多项选择题、填空题、简答题、综合题。

分别为: 单选( $20 \times 1 = 20$  分)      多选( $5 \times 2 = 10$  分)

填空题(每空 1 分共 12 分)      简答( $5 \times 4 = 20$  分)

综合题(4 道共 36 分)

2. 试题中对不同能力层次要求的试题分数实行的比例配置,通常为:识记 20%、理解 40%、应用 40%。

## 第二部分 各章串讲

### 第一章 引 论

#### 内容概要

本章介绍了操作系统的基本概念。主要包括:

计算机系统由硬件和软件两大部分组成,操作系统是计算机系统的一种系统软件,它管理计算机系统的资源和控制程序的执行,改善人机界面和为其他软件提供支持。它的设计目标是使用户方便地使用计算机系统和使得计算机系统能高效地工作。

操作系统的形成和发展与计算机硬件和其他软件的发展密切相关。随着计算机应用的日益广泛,操作系统的功能也日趋完善,根据计算机系统的功能和应用,操作系统可分成几类:批处理操作系统、分时操作系统、实时操作系统、网络操作系统和分布式操作系统。

批处理操作系统按照用户预先规定好的步骤控制作业的执行,实现计算机操作的自动化。批处理多道系统还可以充分利用计算机系统的资源,缩短作业执行时间,提高系统的吞吐率。

分时操作系统支持多个终端用户同时以交互方式使用计算机系统,为用户在测试、修改和控制程序执行方面提供了灵活性。

实时操作系统是实现实时控制的系统,它由外部信号触发而工作,并在特定的时间内完成处理,且给出反馈信号。实时系统对可靠性和安全性的要求极高,不强求系统资源的利用率。

个人计算机系统都是使用微行计算机。比起大型机来,微行机既小又便宜。但是,个人计算机系统的资源和功能相对有限。为了满足较大规模的应用,可把若干台个人计算机系统构成计算机网络。根据计算机网络的结构、通信方式和资源管理方法,分别配置网络操作系统或分布式操作系统。

自考乐园,自考学习交流、资料共享的好去处! 自考乐园,自考人自己的家园...

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

操作系统的资源管理功能可以分为四大部分：处理器管理、存储管理、文件管理、设备管理。

## 重点内容

### （一）操作系统的定义

操作系统是一种管理计算机系统资源、控制程序执行、改善人机界面和为其他软件提供支持的系统软件。

### （二）操作系统的类型

批处理操作系统、分时操作系统、实时操作系统、网络操作系统、分布式操作系统、多机操作系统和嵌入式操作系统。其中前三种是基本的操作系统。

### （三）UNIX 操作系统简介

UNIX 是一个交互式的分时操作系统。UNIX 系统的源代码在它诞生后的初期阶段一直是公开的。

### （四）操作系统的功能

操作系统负责管理计算机系统的所有资源，并调度这些资源的使用。具体来说，其主要功能有：处理器管理、存储管理、设备管理、文件管理。

## 考情分析

本章主要考查几种操作系统的特点，unix 操作系统的介绍，操作系统的用户接口题型一般为单选题，多选题，填空题，有时概念会在简答题中考查。分值大约为 5 分。

例如：09 年 7 月填空题 26 考查了（批处理对作业的控制），09 年 4 月多选题 21 题，09 年 7 月单选题 3 题考查了（实时操作系统的概念和应用），09 年 7 月单选题 2 题考查了（unix 操作系统的产生），09 年 4 月单选题 1 题，填空题 26 题 考查了（操作系统的功能，两类接口）。

## 串讲内容

# 第一节 什么是操作系统（识记）

## 一、计算机系统

1、计算机系统定义：是按用户的要求接收和存储信息、自动进行数据处理并输出结果信息的系统。

2、计算机系统构成：硬件系统和软件系统。

3、硬件系统组成：中央处理器（CPU）、主存储器、辅助存储器、各种输入/输出设备。

## 二、操作系统

1、操作系统定义：是一种管理计算机系统资源、控制程序执行、改善人机界面和为其他软件提供支持的系统软件。

2、操作系统两个主要设计原则：

（1）能使得计算机系统使用方便。→方便

（2）能使得计算机高效地工作。→效率

# 第二节 操作系统的形成（识记）

1、控制台：早期，程序的装入、调试以及控制程序的运行都是程序员通过控制台上的开关来实现。

2、原始汇编系统：用汇编语言编写的程序称为源程序，它不能直接在机器上执行，只有通过汇

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园....

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

编语言解释程序把源程序转换成用机器指令序列表示的目标程序后才能在计算机上运行。

- 3、设备驱动程序：是最原始的操作系统。是一种控制设备工作的程序。
- 4、管理程序：是初级的操作系统。是一种能对计算机硬件和软件进行管理和调度的程序。
- 5、操作系统：采用了 SPOOLING 的处理形式。

SPOOLING 又称“斯普林”。从本质上说，SPOOLING 是把磁盘作为一个巨大的缓冲器。在一个计算问题开始之前，把计算所需要的程序和数据从读卡机或其他输入设备上预先输入到磁盘上存放。这样，在进行计算时不再需要访问读卡机等慢速的输入设备，而可以从速度快得多的磁盘上读取程序和数据。同样，对于计算的结果也是先在磁盘上缓冲存放，待计算完成后，再从打印机上打印出该计算问题的所有计算结果。

### 第三节 操作系统的基本类型（领会）

按照操作系统提供的服务进行分类，可分为批处理操作系统、分时操作系统、实时操作系统、网络操作系统、分布式操作系统、多机操作系统和嵌入式操作系统等。其中批处理操作系统、分时操作系统、实时操作系统是基本的操作系统。

#### 一、批处理操作系统

1、定义：用户为作业准备好程序和数据后，再写一份控制作业执行的说明书。然后把作业说明书连同相应的程序和数据一起交给操作员。操作员将收到的一批作业的有关信息输入到计算机系统中等待处理，由操作系统选择作业，并按其操作说明书的要求自动控制作业的执行。采用这种批量化处理作业的操作系统称为批处理操作系统。

2、分类：

(1) 批处理单道系统：一次只选择一个作业装入计算机系统的主存储器运行。

(2) 批处理多道系统：允许多个作业同时装入主存储器，使中央处理器轮流地执行各个作业，各个作业可以同时使用各自所需要的外围设备。09 年 7 月填空题中考了本知识点。

3、批处理多道系统优点：

(1) 多道作业并行工作减少了处理器的空闲时间，即提高了处理器的利用率。

(2) 作业调度可以按一定的组合选择装入主存储器的作业，只要搭配合理。

(3) 作业执行过程中，不再访问低速的设备，而是直接从高速的磁盘上存取信息，从而缩短了作业执行时间，使单位时间内的处理能力得到提高。

(4) 作业成批输入、自动选择和控制作业执行，减少了人工操作时间和作业交接时间，有利于提高系统的吞吐率。

#### 二、分时操作系统

1、定义：能使用户通过与计算机相连的终端来使用计算机系统，允许多个用户同时与计算机系统进行一系列的交互，并使得每个用户感到好像自己独占一台支持自己请求服务的计算机系统。具有这种功能的操作系统称为分时操作系统，简称分时系统。09 年 7 月的选择题考查了本知识点。

2、分时技术：即把 CPU 时间划分成许多时间片，每个终端用户每次可以使用一个由时间片规定的 CPU 时间。这样，多个终端用户就轮流地使用 CPU 时间。如果某个用户在规定的一个时间片内还没有完成它的全部工作，这时也要把 CPU 让给其他用户，等待下一轮再使用一个时间片的时间，循

环轮转，直至结束。

3、分时系统主要特点：

- (1) 同时性。允许多个终端用户同时使用一个计算机系统。
- (2) 独立性。用户在各自的终端上请求系统服务，彼此独立，互不干扰。
- (3) 及时性。对用户的请求能在较短时间内给出应答。
- (4) 交互性。采用人机对话的方式工作。

### 三、实时操作系统

1、定义：能使计算机系统接收到外部信号后及时进行处理，并且在严格的规定时间内处理结束，再给出反馈信号的操作系统称为实时操作系统，简称为实时系统。

2、实时操作系统的应用十分广泛，如控制科学实验、控制生产流水，监督病人的临床功能、监督和控制飞机的飞行状态，进行工业过程控制等。

3、设计实时系统注意点：

- (1) 要及时响应、快速处理。
- (2) 实时系统要求高可靠性和安全性，不强求系统资源的利用率。

## 第四节 操作系统的发展（领会）

1、单用户微机操作系统：是指早期的微型计算机上运行的操作系统每次只允许一个用户使用计算机。

2、网络操作系统：为计算机网络配置的操作系统称为网络操作系统。网络操作系统把计算机网络中各台计算机系统有机地联合起来，为用户提供一种统一、经济而有效地使用各台计算机系统的方法，可使各台计算机系统相互间传送数据，实现各台计算机系统之间的通信以及网络中各种资源的共享。

3、分布式操作系统：为分布式计算机系统配置的操作系统称为分布式操作系统。分布式操作系统能使系统中若干台计算机相互协作完成一个共同的任务，或者说把一个计算问题可以分成若干个子计算，每个子计算可以在计算机系统内的各计算机上并行执行。

4、多机操作系统：为多处理器系统配置的操作系统称为多机操作系统。

5、嵌入式操作系统：是指运行在嵌入式系统中对各种部件、装置等资源进行统一协调、处理和控制的系统软件。（主要特点是微型化和实时性）

## 第五节 UNIX操作系统简介（识记）

1、诞生：

UNIX 的第一个版本 Version 1 是 AT&T 公司下属的 Bell 实验室里两位程序员 Ken Thompson 和 Dennis Ritchie 凭着个人的兴趣和爱好于 1969 年在一台闲置的 PDP-7 上开发的。

2、特点：

- (1) UNIX 是一个交互式的分时操作系统。
- (2) UNIX 系统的源代码公开。



## 第六节 操作系统的功能（识记）

### 1、操作系统的功能：（从资源管理的角度来分）

- （1）处理器管理：对 CPU 进行管理。
- （2）存储管理：对主存储器进行管理。
- （3）文件管理：通过对磁盘进行管理，实现对软件资源进行管理。
- （4）设备管理：对各类输入/输出设备进行管理。

### 2、操作系统为用户提供的使用接口：

- （1）程序员接口：通过“系统调用”使用操作系统功能。（开发者）
- （2）操作员接口：通过操作控制命令提出控制要求。（应用者）本章介绍了计算机系统的结构和操作系统的结构及相互关系。主要包括：

计算机系统由硬件和软件两大部分组成。硬件是软件执行的基础，硬件具有中央处理器与外围设备并行工作的能力。各个外围设备也可以同时工作。操作系统利用硬件的这种功能，采用多道程序设计技术，允许多用户并行工作。在硬件的中断装置配合下，操作系统能正确地控制各个程序的执行。

对任何一个要执行的程序必须先将它的程序和数据装入主存储器。中央处理器可以直接访问主存储器。各种外围设备只能与主存储器交换信息，在辅助存储器中的信息只有被读入主存储器之后，才能从中央处理器访问。大多数计算机系统把辅助存储器作为主存储器的扩展，用来保护大量的程序和数据。磁盘可以被随机访问，用来存放常用的信息；磁盘主要用于备份，存放不常用的信息以及作为系统间信息交换的媒介。

为了保证计算机系统能正确工作，硬件采用了若干保护措施，操作系统进行配合来提高系统的安全性。中央处理器有两种工作模式：管态（系统态）和目态（用户态）。限定用户程序在目态执行，不能直接使用特权指令。操作系统利用硬件设置的基址寄存器和限长寄存器来限定各程序可访问的主存空间，以免相互干扰而造成错误。

UNIX 是一个交互式的分时操作系统。从结构上看，UNIX 可以分成内核层和外壳层两部分。

操作系统面向用户提供两类使用接口：一组操作控制命令或作业控制语言供用户提出对作业的控制要求；一组系统功能调用为用户程序提供服务功能。

UNIX 允许每一个在终端上登陆成功的用户通过 UNIX 提供的操作控制命令来启动外壳层的用户程序执行。外壳层的用户程序在执行时又可通过系统调用来请求内核层的支持。

由于操作系统十分庞大，所以清晰的结构有利于开发和调试。层次结构是一种重要的设计技术，它不仅使正确性容易得到保证，而且也能提高可维护性和可移植性。

## 第二章 计算机系统结构简介

### 内容概要

本章介绍了计算机系统的结构和操作系统的结构及相互关系。主要包括：

计算机系统由硬件和软件两大部分组成。硬件是软件执行的基础，硬件具有中央处理器与外围设备并行工作的能力。各个外围设备也可以同时工作。操作系统利用硬件的这种功能，采用多道程序设计技术，允许多用户并行工作。在硬件的中断装置配合下，操作系统能正确地控制各个程序的

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园....

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

执行。

对任何一个要执行的程序必须先将它的程序和数据装入主存储器。中央处理器可以直接访问主存储器。各种外围设备只能与主存储器交换信息，在辅助存储器中的信息只有被读入主存储器之后，才能从中央处理器访问。大多数计算机系统把辅助存储器作为主存储器的扩展，用来保护大量的程序和数据。磁盘可以被随机访问，用来存放常用的信息；磁盘主要用于备份，存放不常用的信息以及作为系统间信息交换的媒介。

为了保证计算机系统能正确工作，硬件采用了若干保护措施，操作系统进行配合来提高系统的安全性。中央处理器有两种工作模式：管态（系统态）和目态（用户态）。限定用户程序在目态执行，不能直接使用特权指令。操作系统利用硬件设置的基址寄存器和限长寄存器来限定各程序可访问的主存空间，以免相互干扰而造成错误。

UNIX 是一个交互式的分时操作系统。从结构上看，UNIX 可以分成内核层和外壳层两部分。

操作系统面向用户提供两类使用接口：一组操作控制命令或作业控制语言供用户提出对作业的控制要求；一组系统功能调用为用户程序提供服务功能。

UNIX 允许每一个在终端上登陆成功的用户通过 UNIX 提供的操作控制命令来启动外壳层的用户程序执行。外壳层的用户程序在执行时又可通过系统调用来请求内核层的支持。

由于操作系统十分庞大，所以清晰的结构有利于开发和调试。层次结构是一种重要的设计技术，它不仅使正确性容易得到保证，而且也能提高可维护性和可移植性。

## 重点内容

### （一）计算机系统的结构

#### 1. 计算机系统的层次结构

目前的计算机系统，一般都具有一种层次结构。从内到外依次为：硬件系统、软件系统、支撑软件和应用软件，最外层是使用计算机的人。

#### 2. 操作系统的运行方式

计算机开启时，自动执行引导程序。引导程序首先进行系统初始化的工作，然后把操作系统中的核心装入主存储器。此后操作系统便等待用户请求（事件）的发生，当有某个事件出现，硬件便能识别并能发生一个中断，从而通过操作系统，由它的服务程序去处理，处理结束后，又等待下一个事件发生。

### （二）硬件环境

#### 1. CPU 与外设的并行工作

在计算机系统中，输入/输出控制系统负责完成外围设备与主存储器之间的信息传送。

#### 2. 存储体系

寄存器是处理器的组成部分，用来存放处理器的工作信息。一般来说，寄存器可分成如下几类：

- （1）通用寄存器；
- （2）指令寄存器；
- （3）控制寄存器；

主存储器是 CPU 能直接访问的唯一的存储空间，任何程序和数据都必须被装入主存储器之后，CPU 才能对它进行操作。主存储器以“字节”为单位进行编址，若干字节组成一个“字”。中央处理

器可以按地址读出主存储器中的一个字节或一个字的内容。

为了减少对主存储器的访问时间, 加快程序的执行速度, 现代计算机系统往往配置一个高速缓冲存储器 (cache)。

辅助存储器解决了主存储器的容量不足, 以及主存储器无法永久保存信息的问题。辅助存储器的优点是容量大且能永久保存信息, 缺点是无法被中央处理器直接访问, 必须通过主存储器才能访问。

### 3. 保护措施

在资源共享的计算机系统中, 只有具备了必要的保护措施, 才能使个别的错误不致影响其他程序。硬件的部分保护措施有以下几种:

- (1) 特权指令。
- (2) 管态和目态。
- (3) 存储保护。

## (三) 操作系统的结构

### 1. 操作系统的结构设计目标。

程序的结构是影响程序质量的内在因素。

操作系统结构应追求四大目标, 正确性、高效性、可维护性和可移植性。

### 2. 操作系统的层次结构

层次结构法的最大特点是把整体问题局部化。一个大型复杂的操作系统被分解成若干单向依赖的层次, 由各层的正确性来保证整个操作系统的正确性。

采用层次结构, 能使结构清晰、便于调试, 有利于功能的增加、删除和修改, 正确性容易得到保证, 也提高了系统的可维护性和可移植性。

### 3. UNIX 系统的结构

从结构上看, UNIX 可以分成内核层和外壳层两部分。内核层是 UNIX 操作系统的核心。它具有存储管理、文件管理、设备管理、进程管理等功能, 以及为外壳层提供服务的系统调用。外壳层为用户提供各种操作命令 (UNIX 把它们称为 shell 命令) 和程序设计环境。UNIX 的内核程序用 C 语言和汇编语言编写。

## (四) 操作系统与用户的接口

操作系统为用户提供两类使用接口: 一类是操作员级的, 供用户提出如何控制作业执行的要求; 另一类是程序员级的, 为用户程序提供服务功能。

### 1. 操作控制命令

用户使用作业控制语言编写“作业控制说明书”, 从而形成批处理作业。用户使用键盘输入操作控制命令或选择菜单命令, 进行交互处理。

### 2. 系统调用

操作系统编制了许多不同功能的子程序, 供用户程序在执行中调用, 这些子程序称为系统功能调用程序或系统调用。

## (五) UNIX 的用户接口

UNIX 允许每一个在终端上登陆成功的用户通过 UNIX 提供的操作控制命令来启动外壳层的用户



程序执行。外壳层的用户程序在执行时又可通过系统调用来请求内核层的支持。

(1) shell 命令。

(2) UNIX 系统调用。

### 考情分析

本章主要考查计算机系统的层次结构，操作系统的特权指令、管态与目态操作系统的设计目标，系统调用以及 unix 的系统调用。题型一般式单选题、多选题和简答题。分值大约为 8 分。

例如 09 年 4 月单选题 3 题，09 年 7 月单选题 4 题考查了（特权指令与管态和目态），09 年 4 月简答题考查了（操作系统的设计目标）09 年 4 月单选题考查了（系统调用的概念）

### 串讲内容

## 第一节 计算机系统结构（识记）

### 一、层次结构

#### 1. 计算机系统的层次结构

目前的计算机系统，一般都具有一种层次结构。从内到外依次为：硬件系统、软件系统、支撑软件和应用软件，最外层是使用计算机的人。

硬件系统主要由中央处理器（CPU）、存储器、输入/输出控制系统和各种输入输出/设备组成。

软件系统包括系统软件、应用软件和用用软件组成。系统软禁是计算机系统中与硬件结合最经的软件，也是计算机系统中必不可少的软件。例如，操作系统，编译系统等都是系统软件。

支撑软件是可以支撑其他软件开发和维护的软件，例如，数据库、各种接口软件软件开发工具等都生成软件。

应用软件是按特定领域中的某种需要编写的专用程序，例如。财务管理、人口普查等专用程序都属于应用软件。

2、层次结构：最内层是硬件系统，最外层是使用计算机系统的人，人与硬件系统之间是软件系统。软件系统又依次为系统软件（支撑软件）应用软件。

计算机开启时，自动执行引导程序。引导程序首先进行系统初始化的工作，然后把操作系统中的核心装入主存储器。此后操作系统便等待用户请求（事件）的发生，当有某个事件出现，硬件便能识别并能发生一个中断，从而通过操作系统，由它的服务程序去处理，处理结束后，又等待下一个事件发生。

### 二、系统工作框架

1、引导程序：进行系统初始化，把操作系统中的核心程序装入主存储器，并让操作系统的核心程序占用处理器执行。

2、操作系统核心程序：完成自身的初始工作后开始等待用户从键盘或鼠标输入命令，每接收一条命令就对该命令进行处理。

计算机开启时，自动执行引导程序。引导程序首先进行系统初始化的工作，然后把操作系统中的核心装入主存储器。此后操作系统便等待用户请求（事件）的发生，当有某个事件出现，硬件便能识别并能发生一个中断，从而通过操作系统，由它的服务程序去处理，处理结束后，又等待下一个事件发生。

## 第二节 硬件环境（领会）

### 一、CPU 与外设的并行工作

在现代的通用计算机系统中，为提高计算机的工作效率，均允许中央处理器和外设并行工作。当执行到一条启动外设的指令时，就按指令中给定的参数启动指定的设备，并把控制移交给输入/输出控制系统，由输入/输出控制系统控制外围设备与主存储器之间的信息传送，外围设备独立工作，不再需要中央处理器干预，于是中央处理器可继续执行其他程序。（详见第 6 章）

### 二、存储体系

1、寄存器：是处理器的组成部分，用来存放处理器的工作信息。存取速度快，但造价高。

（1）通用寄存器：存放参加运算的操作数、指令的运算结果等。

（2）指令寄存器：存放当前从主存储器读出的指令。

（3）控制寄存器：存放控制信息以保证程序的正确执行和系统的安全。

A. 程序状态字寄存器：存放当前程序执行时的状态。

B. 中断字寄存器：记录出现的事件。

C. 基址寄存器：设定程序执行时可访问的主存空间的开始地址。

D. 限长寄存器：设定程序执行时可访问的主存空间的长度。

2、主存储器：以字节为单位进行编址。主存储器容量较大，能被处理器直接访问，但断电会丢失数据。

3、高速缓冲存储器：也称 cache，位于处理器和主存储器之间起到缩短存储时间和缓冲存储的作用。

4、辅助存储器：最常用的辅助存储器有磁盘和磁带。优点是容量大且能永久保存信息，但不能被中央处理器直接访问。更多**优质**自考资料尽在百度贴吧**自考乐园**俱乐部

(<http://tieba.baidu.com/club/5346389>) 欢迎♥加入...欢迎♥交流...止不住的惊喜等着你.....

### 三、保护措施

一般是硬件提供保护手段和保护装置，操作系统利用这些设施配合硬件实现保护。

1、指令分类：

（1）特权指令：不允许用户程序中直接执行的指令。如：启动 I/O、设置时钟、设置控制寄存器等。09 年 7 月选择题考了本知识点。

（2）非特权指令：允许用户程序中直接执行的指令。

2、CPU 工作状态：

（1）管态：可执行包括特权指令在内的一切机器指令。一般是操作系统程序占用中央处理器时，CPU 处于管态。

（2）目态：不允许执行特权指令。一般是用户程序占用中央处理器时，CPU 处于目态。09 年 7 月选择题就考了本知识点。

3、存储保护：不同的存储管理方式有不同的实现保护方法（详见第四章）。如可变分区存储管理方式中：基址寄存器的值 $\leq$ 访问地址 $\leq$ 基址寄存器的值+限长寄存器的值。

### 第三节 操作系统结构 (识记)

一、设计目标 09 年 4 月解答题考查了本知识点。

1、正确性：能充分估计和把握各种不确定的情况，使操作系统不仅能保证正确性，且易于验证其正确性。

2、高效性：减少操作系统的开销从而提高计算机系统的效率，尤其对常驻主存储器的核心程序部分更要精心设计。

3、维护性：当系统发现错误或为提高效率而对算法进行调整等工作时，应使操作系统容易维护。

4、移植性：移植性是指能否方便地把操作系统从一个硬件环境移植到另一个新的硬件环境之中。在结构设计时，应尽量减少与硬件直接有关的程序量，且将其独立封装。09 年 4 月选择题考了本知识点。

二、操作系统的层次结构

1、设计方法：无序模块法、内核扩充法、层次结构法、管程设计法等。

2、层次结构法：最大特点是把整体问题局部化。一个大型复杂的操作

系统被分解成若干单向依赖的层次，由各层的正确性来保证整个操作系统的正确性。采用层次结构不仅结构清晰，而且便于调试，有利于功能的增加、删减和修改。

3、操作系统层次结构：处理器管理要对中断事件进行处理，要为程序合理地分配中央处理器的工作事件。它是操作系统的核心程序，是与硬件直接有关的部分，因而把它放在最内层。以后的各层依次存放的是存储管理、设备管理和文件管理。即：硬件→处理器管理→存储管理（设备管理）文件管理。

(1) 层次结构的主要优点是有益于系统的设计和调试。实现各层次的功能时都可利用较内层提供的功能，不必知道内层的功能是怎样实现的，只需知道这些功能可以做什么。调试工作可以从最内层开始，最内层只使用硬件提供的功能，调试时无需考虑其它任何部分。一旦最内层调试通过，则第二层可在最内层提供的功能上执行，同样，调试第三层可在第二层提供的功能上执行，依次类推，知道完成各层功能的调试。若在调试某层的功能时发现错误，可能定错误位于当前层，因为较内层的各层次以通过调试。

(2) 主要困难：层次的划分和安排。由于每一层只能利用位于它内层提供的功能，因而需要仔细地规划每一层的功能，差异保证操作系统工作过程中不出现双向依赖关系。

三、UNIX 系统的结构

1、UNIX 层次结构：内核层和外壳层。

(1) 内核层：是 UNIX 操作系统的核心。它具有存储管理、文件管理、设备管理、进程管理以及为外壳层提供服务的系统调用等功能。

(2) 外壳层：为用户提供各种操作命令和程序设计环境。

2、外壳层组成：由 shell 解释程序、支持程序设计的各种语言、编译程序、解释程序、使用程序和系统库等组成。其中其它模块归 shell 解释程序调用，shell 解释程序用来接收用户输入的命令并进行执行。

3、内核层组成：内核程序用 C 语言和汇编语言编写。按编译方式可分为：汇编语言文件、C 语

言文件和 C 语言全局变量文件。

4、程序运行环境：用户态和核心态。外壳层的程序在用户态运行，内核层的程序在核心态运行。用户态运行的程序称为用户程序，核心态运行的程序称为系统程序。（注：外壳层的用户程序在执行时可通过系统调用来请求内核层的支持）

## 第四节 操作系统与用户的接口（领会）

### 一、操作控制命令

- 1、联机用户：操作控制命令。
- 2、批处理系统用户：作业控制语言，用来编制作业控制说明书。

### 二、系统调用

- 1、系统调用定义：即系统功能调用程序，是指操作系统编制的许多不同功能的供程序执行中调用的子程序。
- 2、执行模式：系统调用在管态下运行，用户程序在目态下运行，用户程序可以通过“访管指令”实现用户程序与系统调用程序之间的转换。（注：访管指令本身是一条在目态下执行的指令）
- 3、系统调用分类：文件操作类、资源申请类、控制类、信息维护类。09 年 4 月解答题考查了本知识点。

## 第五节 UNIX的用户接口（领会）

### 一、shell 命令

#### 1、注册和注销

（1）注册：用户可通过 login 输入用户名和通过 password 输入口令，系统注册成功后在 shell 解释程序控制下，出现提示符（本书采用 C shell 提示符：%）以交互方式为用户服务。

（2）注销：输入 logout 或同时按下 Ctrl+D 键。

#### 2、常见的 shell 命令

- （1）mkdir 命令：请求系统建立一个新的文件目录。
- （2）rmdir 命令：请求系统删除一个空目录。
- （3）cd 命令：切换当前的工作目录。
- （4）pwd 命令：显示用户的当前目录。
- （5）ls 命令：显示一个目录中的文件名。
- （6）cp 命令：复制一个文件。
- （7）mv 命令：对文件重新命名。
- （8）rm 命令：删除一个指定的文件。
- （9）cat 命令：显示用 ASCII 码编写的文本文件。
- （10）more 命令：分屏显示文件内容，按空格键显示下一屏。

#### 3、后台执行的 shell 命令

- （1）方法：在请求后台执行的命令末尾输入字符“&”。
- （2）特点：UNIX 把一个程序转入后台执行后，不等该程序执行完就显示可以输入新命令的提示符。因此，允许多个任务在后台执行，也允许后台任务和前台任务同时执行。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园....

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

#### 4、shell 文件

(1) 定义：用 shell 命令编辑成的文件称为 shell 文件。

(2) 执行 shell 文件：csh shell 文件名

(3) 把 shell 文件改成可执行文件：chmod + x shell 文件名→以后就可直接在提示符后面直接输入文件名就可执行。

## 二、UNIX 系统调用

### 1、常用的系统调用

(1) 有关文件操作的系统调用：

creat：建立文件

open：打开文件

read：读文件

write：写文件

close：关闭文件

link：链接一个文件

unlink：解除文件的链接

lseek：设定文件的读写位置

chmod：改变对文件的访问权限

rename：更改文件名

(2) 有关控制类的系统调用：

fork：创建一个子进程

wait：父进程等待子进程终止

exit：终止子进程的执行

exec：启动执行一个指定文件

(3) 有关信号与时间的系统调用：UNIX 把出现的异常情况或异步事件以传送信号的方式进行。

kill：把信号传送给一个或几个相关的进程

sigaction：声明准备接收信号的类型

sigreturn：从信号返回，继续执行被信号中断的操作

stime：设置日历时间

time：获取日历时间

times：获取执行所花费的时间

2、trap 指令：是 UNIX 系统中的访管指令。

3、系统调用程序入口表：

(1) 作用：实现对系统功能调用程序的统一管理和调度。

(2) 构成：系统调用编号、系统调用所带参数个数、系统调用处理程序入口地址、系统调用名称。

4、系统调用实现过程：

(1) 步骤一：当处理器执行到 trap 指令时便形成一个中断事件。此时将暂停当前用户程序的执行，而由 UNIX 系统内核的“trap 处理子程序”来处理这个中断事件。

(2) 步骤二：trap 处理子程序根据 trap 指令中的系统调用编号查系统调用程序入口表，得到该系统调用所带的参数个数和相应的处理程序的入口地址。然后，把参数传送到内核的系统工作区，再按处理程序入口地址转向该系统调用的处理程序执行。

(3) 步骤三：当系统调用程序完成处理后，仍需返回到 trap 处理子程序，由 trap 处理子程序对被暂停的用户程序进行状态恢复等后续处理，再返回用户程序执行



## 第三章 处理器管理

### 内容概要

本章介绍了处理器管理的基本概念和方法。主要包括：

多道程序设计系统必须做好存储保护、程序浮动及资源分配与调度工作。

多道程序设计系统利用和发挥了处理器与外围设备以及外围设备之间的并行工作能力，从而极大地提高了处理器和其他各种资源的利用率，增加了单位时间内的解题能力。但是，多道程序设计可能会延长程序的执行时间。尤其要注意的是：并行工作道数与系统效率不成正比。

把一个程序在一个数据集合上的一次执行称为一个“进程”。进程是有生命期的，每个进程都有一个进程控制块记录进程的执行情况。随着进程状态的变化，进程经常要从一个队列退出，进入另一个队列，直至进程消亡。

UNIX 中的每一个进程由三部分组成：进程控制块、正文段和数据段。

中断系统在实现进程并发执行、维护系统正常工作、进行故障处理及满足实时处理要求等方面起着重要作用。硬件的中断装置发现并响应中断，操作系统的中断处理程序对出现的中断事件进行处理。一次中断后可能引起若干进程的状态变化，因此，中断处理后应由进程高度决定哪个进程占用处理器。

UNIX 的进程在运行时会由于某些事件的出现而被中断。在 UNIX 中把可能出现的事件分成两大类：中断事件和异常情况。

### 重点内容

#### （一）多道程序设计

让多个计算问题同时装入一个计算机系统的主存储器并行执行，这种程序设计技术称为多道程序设计。

采用多道程序设计技术应注意以下三方面的问题：存储保护；程序浮动；资源的分配和调度。

对具有处理器与外设并行工作能力的计算机系统来说，采用了多道程序设计技术后，能提高整个系统的效率。具体表现为：

- （1）提高了处理器的利用率。
- （2）充分利用外围设备资源。
- （3）发挥了处理器与外围设备以及外围设备之间的并行工作能力。

采用多道程序设计能改善资源使用情况，提高系统效率，但是要注意两个问题：

- （1）可能延长程序执行时间。
- （2）并行工作道数与系统效率不成正比。

#### （二）进程概述

把一个程序在一个数据集上的一次执行称为一个进程。

引入进程的原因有以下两个方面：

- （1）提高资源的利用率。
- （2）正确描述程序的执行情况。

进程有如下一些基本属性：

- (1) 进程的动态性。
- (2) 多个不同的进程可以包含相同的程序。
- (3) 进程可以并发执行。
- (4) 进程有三种基本状态。

为了便于管理进程，我们按进程在执行过程中不同时刻的不同情况定义三种基本状态：

- (1) 等待态—等待某一事件。
- (2) 就绪态—等待系统分配处理器以便运行。
- (3) 运行态—正在占用处理器运行。

根据进程的四个基本属性可概括出进程具有如下三个特性：

1. 动态性：进程是程序的一次执行过程，在执行过程中进程状态不断发生变化。
2. 并发性：若干进程是可同时执行的，它们轮流占用处理器交替执行。
3. 异步性：进程的执行速度取决于自身与外界原因以及能占用处理器的时间，因此以不可预知的速度向前推进。

### (三) 进程队列

为了标识进程，记录各个进程执行时的情况，操作系统在创建进程时为每一个进程设置一个进程控制块。进程控制块的英文名称是 Process Control Block，缩写为 PCB。

在一般情况下，进程控制块应包含四类信息：(1) 标识信息。(2) 说明信息。(3) 现场信息。(4) 管理信息。

操作系统中往往设计一些完成特定功能的，不可中断的过程，这些不可中断的过程称为原语。用于控制进程的原语有：

1. “创建”原语：为一个程序分配一个工作区和建立一个进程控制块，并置该进程为就绪状态。
2. “撤消”原语：一个进程完成工作后，收回它的工作区和进程控制块。
3. “阻塞”原语：进程运行过程中发生等待事件时，把进程状态改为等待态。
4. “唤醒”原语：当进程等待的事件发生时，把进程的状态改为就绪态。

### (四) UNIX 系统中的进程

UNIX 中的进程既可以执行操作系统程序，又可执行用户程序，按照需要进行转换。

由于操作系统要管理系统中各种资源以及进行各种控制，而用户程序只应在操作系统的管理和控制下执行，所以 UNIX 中的进程在用户态或核心态执行时分别拥有不同的权力。

UNIX 中每一个进程由三部分组成：进程控制块、正文段和数据段。其中进程控制块分成两部分：进程基本控制块和进程扩充控制块。正文段是可再入的程序，它由不可被修改的程序和常数组成。UNIX 把进程的数据段划分成三部分：用户栈区、用户数据区和系统工作区。

在 UNIX 系统中，有如下几种状态：(1) 运行状态。(2) 就绪状态。(3) 睡眠状态。(4) 创建状态。(5) 僵死状态。

在 UNIX 中经常要发生进程在主存与磁盘之间的转换，我们把这项工作称为进程的换进换出。标志 runout 和 runin 是交换进程的睡眠标志。

### (五) 中断技术

一个进程占有处理器运行时，由于某些事件的出现使运行被打断，让操作系统处理所出现的事件，到适当的时候再让被打断的进程继续运行，这个过程称为“中断”。引起中断的事件称为“中断源”。对出现的事件进行处理的程序称为“中断处理程序”。在第一章中我们简单地介绍了中断的概念，这里将对中断的问题做进一步的讨论。

### 1. 中断类型

对不同结构的计算机，它们的中断源不尽相同，但从中断事件的性质来说，一般可以分成下述几类：硬件故障中断、程序性中断、外部中断、输入/输出中断、访管中断。

前面四类中断是由于外界的原因迫使正在运行的进程被打断，因此可称为强迫性中断事件，而第五类中断是正在运行的进程所期待的，可称为自愿性中断事件。

### 2. 中断响应

通常，处理器执行完一条指令后，硬件的中断装置立即检查有无中断事件发生。若有中断事件发生，则暂停现行进程的执行，而让操作系统中的中断处理程序占用处理器。这一过程称为“中断响应”。

### 3. 中断处理

中断处理程序对中断事件的处理可分两步进行。第一步是保护好被中断进程的现场信息，即把中断进程的通用寄存器和控制寄存器内容以及被中断进程的 PSW 保存起来，这些信息可以保存在被中断进程的进程控制块中。第二步是对所发生的中断事件进行具体处理。对各类中断事件必须进行不同的处理，中断处理程序分析引起中断的原因后，转交给适当的例行程序来处理该中断。

## (六) UNIX 系统的中断技术

在 UNIX 中把可能出现的事件分成两大类：中断事件和异常情况。

UNIX 用一个由 32 位组成的字作为处理器状态字（记为 ps），可以把它看做 UNIX 中的 PSW。

UNIX 用设置中断优先级的方法起到了中断屏蔽的作用。

在 UNIX 中，对中断事件和异常情况的处理过程十分相似，但有两点主要区别：

(1) 在对异常情况进行处理时，处理器的中断优先级一般是不改变的。而在对中断事件进行处理时，则可以修改处理器的中断优先级。

(2) 对异常情况的处理总是在产生异常情况的那个进程的核心栈上进行。而对中断事件的处理，则离开产生事件的进程而在系统的中断栈上进行。

## (七) 处理器调度

如何对资源进行管理和分配是操作系统中的一个重要问题。处理器调度担负着对处理器的分配工作，它将决定谁能先占用处理器，一次能占用处理器多长时间。

在设计调度算法时，可考虑如下原则：

- (1) 公平性——对用户公平，不能无故或无限制地拖延一个作业的执行。
- (2) 平衡资源使用——尽可能地使系统资源都处于忙碌。
- (3) 极大的流量——在单位时间内为尽可能多的作业服务，保证计算机系统的吞吐能力。

一些常用的作业调度算法：

(1) 先来先服务算法；(2) 计算时间短的作业优先算法；(3) 响应比高者优先算法；(4) 优先级调度算法；(5) 均衡调度算法。

常用的进程调度算法：(1) 先来先服务调度算法；(2) 最高优先级调度算法；(3) 时间片轮转调度算法；(4) 分级调度算法；

UNIX 是一个分时操作系统，但它对进程的调度不是采用时间片轮转法，而是采用了动态优先数调度算法。优先数越小的进程其优先权越高。

### 考情分析

本章主要考查多道程序设计时三方面注意问题，进程的定义，进程的三种状态和其转换，进程控制块的内容，进程的原语，unix 进程控制块的内容，进程状态中断的类型，中断的响应，中断的优先级，作业调度和进程调度的方法。题型有单选题、多选题、填空题、简答题和综合题。

例如 09 年填空题 27 题考查了（多道程序应注意的三方面问题），09 年 4 月简答题 36 题，09 年 7 月简答题 34 题考查了（进程的概念及其和线程的关系），09 年 4 月填空题 22 题，09 年 7 月单选题 7 题考查了（进程控制块的内容）。09 年 4 月单选题考查了（中断的优先级），09 年 4 月综合题 38 题考查了（进程调度的几种方法），09 年 7 月综合题 38 题考察了（作业调度的几种方法）。

### 串讲内容

## 第一节 多道程序设计（领会）

### 一、什么是多道程序设计

1、多道程序设计定义：让多个计算问题同时装入一个计算机系统的主存储器并行执行，这种程序设计技术称为多道程序设计。这种计算机系统称为多道程序设计系统，简称为多道系统。

2、注意事项：

(1) 存储保护：必须提供必要的手段使得在主存储器中的各道程序只能访问自己的区域，避免相互干扰。

(2) 程序浮动：是指程序可以随机地从主存储器的一个区域移动到另一个区域，程序被移动后，仍丝毫不影响它的执行。（可集中分散的空闲区，提高主存空间的利用率）

(3) 资源的分配和调度：多道程序竞争使用处理器和各种资源时，多道程序设计的系统中必须对各种资源按一定的策略进行分配和调度。

### 二、为什么要采用多道程序设计

1、程序的顺序执行：处理器和外围设备，外围设备之间都得不到高效利用。

2、程序的并行执行：让程序的各个模块可独立执行、并行工作，从而发挥处理器与外围设备之间的并行工作能力。

3、多道并行执行：在一个程序各个模块并行工作的基础上，允许多道程序并行执行，进一步提高处理器与外围设备之间的并行工作能力，具体表现在：

(1) 提高了处理器的利用率。

(2) 充分利用外围设备资源。

(3) 发挥了处理器与外围设备以及外围设备之间的并行工作能力。（注：并行：宏观可同时执行，微观也是同时执行。）

### 三、采用多道程序设计应注意的问题

1、可能延长程序执行时间：多道程序设计能提高资源使用效率，增加单位时间的算题量。但是

对每个计算问题来说,从算题开始到全部完成所需的计算时间可能要延长。

2、并行工作道数与系统效率不成正比:并不是并行工作的道数越多,系统的效率就越高,而要根据系统配置的资源 and 用户对资源的要求而定。

- (1) 主存储器空间的大小限制了可同时装入的程序数量。
- (2) 外围设备的数量也是一个制约条件。
- (3) 多个程序同时要求使用同一资源的情况也会经常发生。

## 第二节 进程概述 (领会)

### 一、进程的定义

1、程序:具有独立功能的一组指令或一组语句的集合,或者说是指出处理器执行操作的步骤。  
09年7月解答题中考查了本知识点。

2、进程:是指一个程序在一个数据集上的一次执行。

3、程序和进程的区别:程序是静态的文本,进程是动态的过程。进程包括程序和数据集。

### 二、为什么要引入进程

1、提高资源的利用率:一个程序被分成若干个可独立执行的程序模块,每个可独立执行的程序模块的一次执行都可看做一个进程,通过进程的同步可提高资源的利用率。

2、正确描述程序的执行情况:可以方便描述一个程序被执行多次时,各自的执行进度。

### 三、进程的属性

进程有如下一些基本属性:

- (1) 进程的动态性。
- (2) 多个不同的进程可以包含相同的程序。
- (3) 进程可以并发执行。
- (4) 进程有三种基本状态。

为了便于管理进程,我们按进程在执行过程中不同时刻的不同情况定义三种基本状态:

- (1) 等待态—等待某一事件。
- (2) 就绪态—等待系统分配处理器以便运行。
- (3) 运行态—正在占用处理器运行。

进程状态变化的几种情况有:

- (1) 运行态→等待态
- (2) 等待态→就绪态
- (3) 运行态→就绪态
- (4) 就绪态→运行态

根据进程的四个基本属性可概括出进程具有如下三个特性:

- 1. 动态性:进程是程序的一次执行过程,在执行过程中进程状态不断发生变化。
- 2. 并发性:若干进程是可同时执行的,它们轮流占用处理器交替执行。
- 3. 异步性:进程的执行速度取决于自身与外界原因以及能占用处理器的时间,因此以不可预知的速度向前推进。



### 第三节 进程队列（领会）

#### 一、进程控制块

- 1、进程控制块作用：即 PCB，是进程存在的标识。
- 2、进程控制块构成：
  - （1）标识信息：用来标识进程的存在和区分各个进程。进程名。
  - （2）说明信息：用于说明本进程的情况。包括：进程状态、等待原因、进程程序存放位置、进程数据存放位置。
  - （3）现场信息：用来当进程由于某种原因让出处理器时，记录与处理器有关的各种现场信息。包括：通用寄存器内容、控制寄存器内容、程序状态字寄存器内容。
  - （4）管理信息：用来对进程进行管理和调度的信息。包括：进程优先级、队列指针。

#### 二、进程的创建和撤销

- 1、进程创建：当系统为一个程序分配一个工作区（存放程序处理的数据集）和建立一个进程控制块后就创建了一个进程。刚创建的进程其状态为就绪状态（若执行过程中还缺少资源可以再将其转为等待状态）。
- 2、进程的撤销：当一个进程完成了特定的任务后，系统收回这个进程所占的工作区和取消该进程的进程控制块，就撤销了该进程。
- 3、原语：是操作系统设计用来完成特定功能且不可中断的过程。包括：创建原语、撤销原语、阻塞原语、唤醒原语。
  - 1.“创建”原语：为一个程序分配一个工作区和建立一个进程控制块，并置该进程为就绪状态。
  - 2.“撤消”原语：一个进程完成工作后，收回它的工作区和进程控制块。
  - 3.“阻塞”原语：进程运行过程中发生等待事件时，把进程状态改为等待态。
  - 4.“唤醒”原语：当进程等待的事件发生时，把进程的状态改为就绪态。

#### 三、进程队列的链接

- 1、进程队列概念：为了便于管理，经常把处于相同状态的进程链接在一起，称为进程队列。
- 2、进程队列分类：
  - （1）就绪队列：是指把若干个等待运行的进程（就绪进程）按一定的次序链接起来的队列。
  - （2）等待队列：是指把若干个等待资源或等待某些事件的进程按一定的次序链接起来的队列。
- 3、队列实现方法：只需将状态相同的进程控制块链接起来就可以。链接的方式包括单向链接和双向链接。
- 4、队列管理：是指系统中负责进程入队和出队的工作。
  - （1）入队：是指一个进程进入到指定的队列。
    - A. 从队首入队成为新的队首进程。
    - B. 从队尾入队成为新的队尾进程。
    - C. 插入到队列中某两个进程之间。
  - （2）出队：是指一个进程从所在的队列中退出。也存在三种情况。

## 第四节 UNIX系统中的进程（领会）

### 一、UNIX 进程的特点

UNIX 中的进程执行用户程序时在用户态执行，执行操作系统程序时在核心态执行。在用户态执行的进程请求系统功能调用时，便转换到核心态执行操作系统程序，当一次系统调用结束时，该进程从核心态的执行返回到用户态执行用户程序。

### 二、UNIX 进程的组成

#### 1、进程控制块：

(1) 进程基本控制块：用来记录进程调度时必须使用的一些信息，常驻主存储器。把进程基本控制块的数据结构称为 proc 结构。

A. 标识信息：包括用户标识（分为实际用户标识号和设置用户标识号）和进程标识。

B. 有关进程非常驻主存部分的信息：用来建立信息在主存与磁盘之间传送。包括：非常驻主存部分所在的地址、长度和一些必要的指针。

C. 有关进程调度的信息：包括：进程状态、标志、优先数以及调度有关的其他信息。

D. 其他信息：用于管理和控制的信息，如进程扩充控制块的地址、进程共享正文段和共享主存段的管理信息、进程接收的信号。

(2) 进程扩充控制块：随用户程序和数据装入主存储器或调出主存储器。把进程扩充控制块的数据结构称为 user 结构。包括：标识、现场保护、主存管理、文件读写、系统调用、进程控制与管理等。

2、正文段：是指 UNIX 中可供多个进程共享的程序。系统中设置了一张正文表 text[]，用来指出该正文段在主存和磁盘上的位置、段的大小和调用该正文段的进程数等情况。

3、数据段：包括进程执行的非共享程序和程序执行时用到的数据。

(1) 用户栈区：是进程在用户态执行时的工作区，主要用于函数调用的参数传递、现场保护、存放返回地址、存放局部变量等。

(2) 用户数据区：存放进程执行中的非共享程序和用户工作数据。

(3) 系统工作区：

A. 核心栈：是进程在核心态执行时的工作区，主要用于函数调用的参数传递、现场保护、存放返回地址、存放局部变量等。

B. user 区：存放进程扩充控制块。

### 三、UNIX 进程的状态

UNIX 进程的状态有运行状态、就绪状态、睡眠状态、创建状态、僵死状态。09 年 4 月解答题考查了本知识点。

在 UNIX 中经常要发生进程在主存与磁盘之间的转换，我们把这项工作称为进程的换进换出。标志 runout 和 runin 是交换进程的睡眠标志。

### 四、UNIX 进程的创建和终止

1、UNIX 的进程树：0 号进程（也称交换进程，是系统启动后 UNIX 的核心程序完成初始化后创建的第一个进程，在核心态运行，用来进行进程调度和让进程在主存与磁盘上进行交换）→1 号进

程（也称初始化进程，由 0 号进程创建，在用户态运行，用来为终端用户请求注册时创建 login 进程）→login 进程（用来处理用户的登录过程，登录成功后创建 shell 进程）→shell 进程（等待用户输入命令）。

2、进程的创建：在 UNIX 中，除了 0 号进程和 1 号进程外，其他的进程总是使用系统调用 fork 来创建新进程，形成父子关系。子进程是父进程的一个映像，除了进程的状态、标识和时间有关的控制项外，全部复制父进程的 proc、user、栈和数据区。

(1) fork 的主要工作

(2) 实现子进程可与父进程执行不同的程序段

3、进程的终止：

(1) 系统调用 exit 的主要任务是把终止进程自被创建以来所占用的系统资源退还给系统。关闭该进程所有打开的文件，释放它对正文段的使用权，把它的 user 结构换出到磁盘对换区后收回数据段占用的主存空间。此后，把终止进程的状态改为“僵死状态”，向父进程发出信号，由父进程作善后处理。

(2) 系统调用 wait 要对用 exit 请求终止的进程作善后处理。当进程用系统调用 wait 等待其子进程终止时，wait 的任务是先查找处于僵死状态的子进程，若子进程尚未僵死，则让该进程等待，直到子进程成为僵死状态后被释放。进程被释放后，wait 继续执行，再从磁盘对换区中把该子进程的 user 结构读入主存缓冲区，释放该 user 在对换区所占的空间。然后，把保存在 user 中的子进程的时间信息加入到本进程的 user 结构中，再释放主存缓冲区，把子进程在 proc[ ] 中的表项清除。

## 五、UNIX 进程的换进换出

在 UNIX 中经常要发生进程在主存与磁盘之间的转换，我们把这项工作称为进程的换进换出。这项工作由交换进程（即 0 号进程）执行 sched 程序来完成。标志 runout 和 runin 是交换进程的睡眠标志。当磁盘对换区中没有要换进的进程时，标志 runout 置为“1”，交换进程睡眠，直到对换区有要换进的进程时被唤醒。当磁盘对换区有就绪进程要换进，但没有足够的主存空间，也没有可换出的进程，则标志 runin 置为“1”，交换进程睡眠，直到主存有进程可换出时被唤醒。

## 六、UNIX 进程的睡眠与唤醒

1、进程的睡眠：一般说，进程总是在执行一个系统调用时被确定是否应睡眠。所以，进程的状态也总是从“核心态运行”变成“在主存睡眠”。

在确定一个进程需睡眠时，便调用 sleep 程序让进程进入睡眠状态，且将其链入睡眠队列。2、进程的唤醒：通过调用 wakeup 程序来唤醒等待相应事件的进程。被唤醒的进程从睡眠队列退出，状态被修改为就绪，在主存睡眠而被唤醒的进程链入在主存的就绪队列，在磁盘对换区睡眠的进程被唤醒后仍保留在对换区，链入就绪且换出队列。

# 第五节 中断技术（领会）

## 一、中断和中断类型

1、中断基本概念：一个进程占有处理器运行时，由于自身或外界的原因使运行被打断，让操作系统处理所出现的事件，到适当的时候再让被打断的进程继续运行，我们称该进程被中断了。引起中断的事件称为中断源。对出现的事件进行处理的程序称为中断处理程序。

2、中断类型：按中断事件的性质来分。

(1) 强迫性中断事件：这类中断事件不是正在运行的进程所期待的，而是由于外部的请求或某些意外事故而使正在运行的进程被打断。

A. 硬件故障中断：由计算机故障造成的中断，如电源电压超出规定范围。

B. 程序性中断事件：由执行到程序的某条指令出现的问题引起的中断，如除数为“0”。

C. 外部中断事件：由各种外部事件引起的中断，如用户从终端上输入了一条命令。

D. 输入/输出中断事件：由来自输入/输出控制系统的事件所引起的中断，如外围设备完成了一次信息传输操作。

(2) 自愿性中断事件：这是正在运行的进程所期望的中断事件，是正在运行的进程执行一条“访管指令”请求系统调用为其服务所引起的中断，也称为访管中断。

## 二、中断响应

1、中断响应概念：若有中断事件发生，则暂停现行进程的执行，而让操作系统的中断处理程序占用处理器，这一过程称为中断响应。

2、中断装置的具体工作：

(1) 检查是否有中断事件发生。

(2) 若有中断事件发生，则暂停现行进程的执行，且保护好被中断进程的断点以及其他一些信息，以便进程在适当的时候能继续执行。

(3) 启动操作系统的中断处理程序工作。

3、中断字寄存器：用来记录强迫性中断事件的寄存器。中断装置只要检查中断字寄存器就可知道是否有中断事件发生。若中断字寄存器的内容为“0”，则无中断事件发生，处理器继续执行下一条指令；若中断字寄存器的内容为非“0”，则表示有中断事件发生。(自愿性中断由执行访管指令时分析操作码可知)

4、程序状态字和程序状态字寄存器：

(1) 程序状态字 (PSW)：用来控制指令执行顺序并且保留和指示与程序有关的系统状态。

(2) 程序状态字寄存器：在计算机系统中，对每个处理器设置一个用来存放当前运行进程的 PSW 的寄存器，该寄存器称为程序状态字寄存器。

5、中断响应：

(1) 当前 PSW：存放在程序状态字寄存器中的 PSW 是当前正在占用处理器的进程的 PSW。

(2) 新 PSW：出现中断事件后，要由操作系统的中断处理程序占用处理器，让中断处理程序处理出现的中断事件。我们把中断处理程序的 PSW 称为新 PSW。

(3) 旧 PSW：中断处理程序在占用处理器前，必须把被中断进程的 PSW 保护好，以便该进程在适当的时候按被中断时的情况继续执行。我们把保护好的被中断进程的 PSW 称为旧 PSW。

## 三、中断事件的处理

1、保护被中断进程的现场信息：把中断时的通用寄存器内容、控制寄存器内容以及已被中断装置保存的旧 PSW 保存到被中断进程的进程控制块中。

2、分析中断原因：根据被中断时由中断装置保存的旧 PSW 中的中断码可知发生该种中断的具体原因。

3、处理发生的中断事件：在多数情况下中断处理程序只需做一些保护现场、分析事件性质等原则性的处理，而具体的处理可由适当的例行程序来完成。

#### 四、中断优先级和中断屏蔽

1、中断优先级：一般说来，中断装置是按预定的顺序响应同时出现的中断事件，这个预定的顺序称为中断优先级。一般情况下，优先级的高低顺序依次为硬件故障中断、自愿中断、程序性中断、外部中断、输入/输出中断。

2、中断屏蔽：是指中断发生后，若出现优先级别比正在处理的中断级别高是否可再中断，形成中断的嵌套。一般中断屏蔽只屏蔽比自己级别低的中断事件。另外，资源中断是不能屏蔽的。

### 第六节 UNIX系统的中断技术（识记）

#### 一、中断事件和异常情况

1、中断事件：如果出现的事件与正在运行的进程无关，则把这些事件称为中断事件。如 I/O 中断事件、时钟中断事件等。

2、异常情况：如果出现的事件与正在运行的进程有关，则把这些事件称为异常情况。如执行到 trap 指令、地址越界等。

#### 二、处理器状态字

UNIX 用一个由 32 位组成的字作为处理器状态字（记为 ps）。16-31 位的内容为特权信息，只允许内核程序对它作修改；0-15 位的内容为非特权信息，允许任何一个程序按规定的含义自由设置和修改。

#### 三、中断处理

##### 1、中断响应：

（1）把程序计数器寄存器中的内容（旧 pc）和处理器状态字寄存器中的内容（旧 ps）保存在指定的内部寄存器中。

（2）按出现的事件查处理程序入口表，从中取出处理程序的入口地址（新 pc），把它送到程序计数器寄存器中。

（3）把取到的处理器状态字（新 ps）送到处理器状态字寄存器中，从而操作系统的处理程序就占用了处理器。

##### 2、中断处理过程：

（1）处理过程：现场保护 ◇ 分析处理 ◇ 恢复现场。

（2）中断事件与异常情况处理过程的区别：

A. 在对异常情况进行处理时，处理器的中断优先级一般是不改变的。而在对中断事件进行处理时，则可以修改处理器的中断优先级。

B. 对异常情况的处理总是在产生异常情况的那个进程的核心栈上进行。而对中断事件的处理，则离开产生事件的进程而在系统的中断栈上进行。

### 第七节 处理器调度（综合应用）

#### 一、处理器的两级调度

1、作业调度：是指从输入井中选取后备作业装入主存储器的工作。（作业调度应遵循的必要条件  
自考乐园, 自考学习交流、资料共享的好去处！ 自考乐园, 自考人自己的家园...

俱乐部 id: **5346389**（请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部



件：系统现有的尚未分配的资源可以满足被选作业的资源要求。)

2、进程调度：是指从就绪进程中选取一个进程，让它占用处理器的工作。→作业调度和进程调度的关系：任何一个作业，只有先被作业调度选中才有机会去竞争处理器，然后仅当被进程调度选中时才能占用处理器。

## 二、批处理作业的调度算法

1、设计调度算法原则 09 年 4 月选择题就考查本知识点。

- (1) 公平性：对用户公平，不能无故或无限制地拖延一个作业的执行。
- (2) 平衡资源使用：尽可能地使系统都处于忙碌。
- (3) 极大的流量：在单位时间内为尽可能多的作业服务，保证计算机系统的吞吐能力。

2、周转时间：假定作业  $i$  进入输入井的时间为  $S_i$ 。若它被选中执行，得到计算结果的时间为  $E_i$ ，那么它的周转时间就定义为  $T_i = E_i - S_i$ 。

3、作业调度算法：

- (1) 先来先服务算法
- (2) 计算时间短的作业优先算法
- (3) 响应比高者优先算法：响应比 = 等待时间 / 计算时间。
- (4) 优先级调度算法
- (5) 均衡调度算法

## 三、进程调度算法

1、进程切换：是指一个进程让出处理器由另一个进程占用处理器的过程。引起进程切换的事件：

- (1) 一个进程从运行状态变成等待状态。
- (2) 一个进程从运行状态变成就绪状态。
- (3) 一个进程从等待状态变成就绪状态。
- (4) 一个进程完成工作后被撤销。

2、进程调度算法：

- (1) 先来先服务调度算法
- (2) 最高优先级调度算法
- (3) 时间片轮转调度算法
- (4) 分级调度算法：要求掌握调度的原则。

## 四、UNIX 系统的进程调度算法

1、进程调度算法：UNIX 采用动态优先数调度算法。UNIX 中每个进程都有一个优先数，进程的优先数随进程的执行情况而变化。就绪进程能占用处理器的优先权取决于进程的优先数，优先数越小则优先权越高。

2、进程优先权确定原则：

- (1) 进入核心态运行的进程优先权高于在用户态运行的进程优先权。
- (2) 一个进程因用完了一个时间片而被剥夺处理器时，应降低该进程的优先权，以使其他进程有机会使用处理器。
- (3) 对进入睡眠的进程，系统将按照它们等待事件的轻重缓急赋予它们不同的优先权。

(4) 应相应降低累计使用处理器时间较长的进程的优先权,以减少这些进程找那用处理器的机会。

### 3、进程的优先数确定方法:

(1) 设置法: 用于即将进入睡眠状态的进程,当进程由于某个事件要睡眠时,根据事件的性质对该进程设置优先数。对因紧迫的事件而入睡的进程设置较小的优先数,一般为负值(其优先权就高);对等待的事件不会产生全局性影响的睡眠进程设置较大的优先数,一般为正值。

(2) 计算法: 当进程转入用户态运行时,则用计算法来确定其优先数。系统对正在占用处理器的进程每隔一个时钟周期,就会变大其优先数,优先权就相应降低,在进程切换时再次占用处理器的机会就减少。

### 4、进程调度程序 swtch:

#### (1) 引起进程调度的原因:

- A. 进程完成了预定的工作而终止。
- B. 进程因等待某些事件而进入睡眠状态。
- C. 进程用完了一个规定的时间片。
- D. 对捕获到的异常情况处理结束后。

(2) swtch 主要任务: 在主存就绪的进程中,选择一个优先数最小的进程;为被选中的进程恢复现场信息。

## 第四章 存储管理

### 内容概要

本章介绍了计算机系统存储器管理的基本概念和方法。主要包括:

重定位的概念、固定分区存储管理、可变分区存储管理、页式存储管理、单用户连续存储管理和虚拟存储器。

主存储器是可被处理器直接访问的,处理器是按绝对地址访问主存储器的。为了使用户编制的程序能存放在主存储器的任意区域执行,用户使用逻辑地址编辑程序,即用户使用了逻辑上的主存储器。存储管理必须为用户分配一个物理上的主存空间,于是就有一个从逻辑空间到物理空间的转换问题。具体地说,是要把逻辑地址转换成绝对地址,把这样的地址转换工作称为重定位。

存储管理必须合理地分配主存空间,为了避免主存中的各程序相互干扰还必须实现存储保护,为了有效利用主存空间允许多个作业共享程序和数据,各种存储管理方式实现这些功能的方法是不同的,并且都有相应的硬件做支持。

实现虚拟存储器后,从系统的角度看提高了主存空间的利用率,从用户的角度看,编制程序不受主存实际容量的限制。虚拟存储器的容量由地址结构决定,若地址用  $n$  位表示,则虚拟存储器的最的容量为  $2^n$ 。虚拟存储器的实现借助于大容量的辅助存储器(例如磁盘)存放虚存中的实际信息,操作系统利用程序执行在时间上和空间上的局部性特点把当前需要的程序段和数据装入主存储器,且利用表格(例如页表、段表)构造一个用户的虚拟空间。硬件根据建立的表格进行地址转换或发出需进行调度的中断信号(例如缺页中断、缺段中断等)。操作系统处理这些中断事件时,选择一种合适的调度算法对主存储器和辅助存储器中的信息进行调出和装入,尽可能地避免“抖动”。

UNIX 采用页式虚拟存储管理，并把编程序时用的地址称为虚拟地址。

## 重点内容

### (一) 概述

现代计算机系统都采用多级存储体系。这个存储体系主要包括主存储器、辅助存储器、高速缓冲存储器和寄存器。

为了提高处理器的利用率和系统的工作效率，主存储器中经常存放多个程序和数据。许多计算机系统都采用二级存储的办法，利用辅助存储器提供的大容量存储空间，存放准备运行的程序和数据，当需要时或主存空间允许时，随时将它们读入主存储器。

存储管理是对主存空间的用户区进行管理，其目的是尽可能地方便用户和提高主存空间的利用率。具体地说，存储管理的功能如下：

1. 主存空间的分配与回收。
2. 实现地址转换。
3. 主存空间的共享与保护。
4. 主存空间的扩充。

### (二) 重定位

绝对地址：主存空间的地址编号。与绝对地址对应的主存空间称为物理地址空间。

相对地址：用户程序中使用的地址。与相对地址对应的存储空间称为逻辑地址空间。

为了保证作业的正确执行，必须根据分配给作业的主存空间对作业中指令和数据的存放地址进行转换，即要把相对地址转换成绝对地址。把相对地址转换成绝对地址的工作称为重定位或地址转换。

重定位的方式可以有静态重定位和动态重定位两种。

### (三) 单用户连续存储管理

这是一种最简单的存储管理方式，系统是将整个主存空间除了给操作系统划分出一块空间外，其余部分的空间都分配给一个作业使用。个人计算机可采用此种管理方法，它不适宜多道程序设计系统。

如果作业的逻辑地址空间大于用户区，采用覆盖技术可以把作业装入主存储器中。

某些系统还采用对换技术让多个进程轮流进入内存，这种技术多用于分时系统，随着进程调度，将内存中的进程暂时移到外存，而把外存中某一进程换进内存。

### (四) 固定分区存储管理

其基本思想是将内存划分成若干个固定大小的分区，每个分区中最多只能装入一个作业。当作业申请内存时，系统按一定的算法为其选择一个适当的分区，并装入内存运行。由于分区大小是事先固定的，因而可容纳作业的大小受到限制，而且当用户作业的地址空间小于分区的存储空间时，容易造成存储空间浪费。

1. 主存空间的分配与回收

系统设置一张“分区分配表”来描述各分区的使用情况。登记的内容应包括：分区号、起始地址、长度和占用标志。其中占用标志为“0”时，表示目前该分区空闲；否则登记占用作业名（或作业号）。有了“分区分配表”，空间分配与回收工作是比较简单的。

## 2. 地址转换和存储保护

固定分区管理可以采用静态重定位方式进行地址映射。

为了实现存储保护，处理器设置了一对“下限寄存器”和“上限寄存器”。当一个已经装入主存储器的作业能够得到处理器运行时，进程调度应记录当前运行作业所在的分区号，且把该分区的下限地址和上限地址分别送入下限寄存器和上限寄存器。处理器执行该作业的指令时必须核对其要访问的绝对地址是否越界。

### (五) 可变分区存储管理

可变分区存储管理不是预先将主存划分分区，而是在作业转入主存时建立分区，使分区的大小正好与作业要求的存储空间相等。这种处理方式使内存分配有较大的灵活性，也提高了内存利用率。但是随着对内存不断地分配、释放操作会引起存储碎片的产生。

#### 1. 主存空间的分配与回收

采用可变分区存储管理，系统中的分区个数与分区的大小都在不断地变化，系统利用“空闲区表”来管理内存中的空闲分区，其中登记空闲区的起始地址、长度和状态。当有作业要进入内存时，在“空闲区表”中查找状态为“未分配”且长度大小或等于作业的空闲分区分配给作业，并做适当调整；当一个作业运行完成时，应将该作业占用的空间作为空闲区归还给系统。

可以采用最先适应算法、最优适应算法和最坏适应算法三种分配策略之一进行内存分配。

#### 2. 地址转换和存储保护

可变分区存储管理一般采用动态重定位的方式。为实现地址重定位和存储保护，系统设置了相应的硬件：基址/限长寄存器（或上界/下界寄存器）、加法器、比较线路等。

基址寄存器用来存放程序在内存的起始地址，限长寄存器用来存放程序的长度。处理器在执行时，用程序中的相对地址加上基址寄存器中的基地址，形成一个绝对地址，并将相对地址与限长寄存器进行计算比较，检查是否发生地址越界。

#### 3. 移动技术

所谓移动是指把作业从一个存储区域移到另一个存储区域的工作。采用移动技术有两个目的：

- (1) 集中分散的空闲区；
- (2) 便于作业动态扩充主存。

移动可集中分散的空闲区，提高主存空间的利用率。移动也为作业动态扩充主存空间提供了方便。但是采用移动技术时必须注意下列问题：

- (1) 移动会增加系统开销；
- (2) 移动是有条件的。

所以，在采用移动技术的系统中，应尽可能地减少移动，以降低系统开销，提高系统效率。为此可改变作业装入主存储器的方式来达到减少移动的目的。

### (六) 页式存储管理

#### 1. 基本原理

页式存储管理是把主存储器分成大小相等的许多区，每个区称为一块。采用分页式存储管理时，相对地址由两部分组成：页号和页内地址。在进行存储空间分配时，总是以块为单位进行分配。

页式存储管理必须解决两个关键的问题：第一，怎样知道主存储器上哪些块已被占用，哪些块

是空闲的；第二，作业信息被分散存放后如何保证作业的正确执行。

## 2. 页式主存空间的分配与回收

采用页式存储管理，可用一张“位示图”来构成主存分配表。进行主存分配时，先查空闲块数能否满足作业要求。若不能满足，则作业不能装入。作业执行结束后，应收回作业所占的主存块。

## 3. 页表和地址转换

当主存中空闲块数能满足作业要求时，存储管理就找出这些空闲块分配给作业，同时为作业建立一张页表，指出相对地址中页号与主存中块号的对应关系。

页式存储管理采用动态重定位的方式装入作业，作业执行时由硬件的地址转换机构来完成从相对地址到绝对地址的转换工作。页表是进行地址转换的依据。

利用高速缓冲存储器存放页表的一部分，把存放在高速缓冲存储器中的部分页表称为快表。采用快表后，使得地址转换的时间大大下降。

## （七）虚拟存储管理

虚拟存储器实际上是为扩大主存容量而采用的一种管理技巧。采用虚拟存储有两个好处，第一，使主存空间能充分地被利用；第二，从用户的角度来看，好象计算机系统提供了容量很大的主存储器。

虚拟存储器的工作原理：把作业信息保留在磁盘上，当要求装入时，只将其中一部分先装入主存储器，作业执行过程中，若要访问的信息不在主存中，则再设法把这些信息装入主存。

页式虚拟存储器的实现是借助大容量辅助存储器（如磁盘）存放虚存中的实际信息，操作系统利用程序执行时在时间上和空间上的局部性特点，把当前需要用的程序段和数据装入主存储器。且利用表格（如页表）为用户构造一个虚拟空间（对不在主存的信息，指出存放在辅助存储器中的位置）。硬件根据操作系统建立的表格进行地址转换，当发现所要访问的信息不在主存储器中时，发出需要调度的中断信号（如缺页中断）。操作系统处理这个中断事件时，选择一种好的调度算法对主存储器和辅助存储器中的信息进行高效调度，尽可能避免“抖动”。页面调度算法有最佳调度算法、先进先出调度算法、最近最久未使用调度算法、最近最不经常使用调度算法等。

## （八）UNIX 系统的页式虚拟存储管理

UNIX 采用页式虚拟存储管理，并把编程序时用的地址称为虚拟地址。

UNIX 把进程的地址空间分成三个区段，它们是系统区段、程序区段和控制区段。

UNIX 系统页面调度使用二次机会替换算法，尽量减少输入/输出的传送次数，以提高系统的效率。

## 考情分析

本章主要考查绝对地址和相对地址的转换，对换技术，主存空间的分配与回收，地址转换和存储保护，页式管理的原理及方法，页表和地址转换，页面调度的方法 unix 页表和地址转换，题型一般为单选题，多选题，填空题，综合题。分值大约为 20 分。

例如：09 年 4 月单选题 9 考查了（绝对地址和相对地址的转换），09 年 4 月单选题 10 题考查了（对换技术），09 年 7 月单选题 12 题考查了（虚拟存储的保护），09 年 4 月单选题 11 题考查了（页式管理的原理），09 年 4 月填空题 29 考查了（页表的地址转换）09 年 4 月综合题 39 考查了（页面调度的几种算法）



## 串讲内容

### 第一节 概述 (识记)

#### 一、信息的二级存储

利用辅助存储器提供的大容量存储空间,存放准备运行的程序和数据,当需要时或主存空间允许时,随时将它们读入主存储器。

#### 二、存储管理的功能

##### 1、主存储器空间划分:

(1) 系统区:用来存放操作系统与硬件的接口信息、操作系统的管理信息和程序、标准子程序等。

(2) 用户区:用来存放用户的程序和数据。

存储管理是对主存空间的用户区进行管理,其目的是尽可能地方便用户和提高主存空间的利用率。具体地说,存储管理的功能如下:

1. 主存空间的分配与回收。
2. 实现地址转换。
3. 主存空间的共享与保护。

##### 4. 主存空间的扩充。更多**优质**自考资料尽在百度贴吧**自考乐园**俱乐部

(<http://tieba.baidu.com/club/5346389>) 欢迎♥加入...欢迎♥交流...止不住的惊喜等着你.....

### 第二节 重定位 (领会)

#### 四、绝对地址和相对地址 09 年 4 月选择题就考了本知识点。

绝对地址:主存空间的地址编号。与绝对地址对应的主存空间称为物理地址空间。

相对地址:用户程序中使用的地址。与相对地址对应的存储空间称为逻辑地址空间。(用户作业都从“0”地址开始往下编写) 09 年四月选择题就考察了本知识点。

为了保证作业的正确执行,必须根据分配给作业的主存空间对作业中指令和数据的存放地址进行转换,即要把相对地址转换成绝对地址。把相对地址转换成绝对地址的工作称为重定位或地址转换。

重定位的方式可以有静态重定位和动态重定位两种

#### 二、重定位的方式

1、重定位:也称地址转换,是指把相对地址转换成绝对地址的工作。(分为静态重定位和动态重定位)

2、静态重定位:地址转换的时刻在入主存储器时。

3、动态重定位:地址转换的时刻在入处理器时。

→动态重定位支持程序浮动,即作业执行时,被改变了存放区域的作业仍然能正确执行。而采用静态重定位时,由于装入主存储器的作业信息已经都是用绝对地址指示,故作业在执行过程中时不能移动位置的。

### 第三节 单用户连续存储管理（领会）

#### 一、存储空间的分配

1、存储空间分配方法：处理器中设置一个界限寄存器，存放当前可供用户使用的主存区域的起始地址。作业入主存储器时总是存放到由界限寄存器指示的起始地址开始往下存储。

2、重定位：采用静态重定位方式进行地址转换，即在作业装入主存时，由装入程序完成地址转换。装入程序只要把界限寄存器的值加到相对地址上就可完成地址转换。

3、存储保护：主存最大地址 $\geq$ 绝对地址 $\geq$ 界限地址，成立则可执行，否则有地址错误，形成“地址越界”的程序性中断事件。

4、缺点：

（1）当作业执行中出现了某个等待事件时，处理器就处于空闲状态，不能被利用。

（2）一个作业独占主存中的用户区，当主存中有空闲区域时，也不能被其他作业利用，降低了主存空间的利用率。

（3）外围设备也不能充分被利用。

#### 二、覆盖技术

1、将作业划分成若干段，其中有一个主段是作业执行过程中经常要用到的信息，而其他段是不会同时工作的。

2、主段入驻留区；其他段轮流入覆盖区。

#### 三、对换技术

在分时系统中，单用户连续存储管理可用对换方式让多个用户的作业轮流进入主存储器执行。系统中必须要有一个大容量的高速辅助存储器，多个用户的作业信息都被保留在磁盘上，把一个作业先装入主存储器让它执行。当执行中出现等待事件或用完一个时间片时，把该作业从存储器换出，再把由调度程序选中的另一个作业换入到主存储器中。09年4月选择题考查了本知识点。

### 第四节 固定分区存储管理（领会）

#### 一、基本原理

将主存储器中可分配的用户区域预先划分成若干个连续区，每个连续区成为一个分区。每个分区的大小可以相同，也可以不同。每个分区可用来装入一个作业，但不允许在一个分区中同时装入多个作业。

#### 二、主存空间的分配与回收

系统设置一张“分区分配表”，用来说明各分区的分配和使用情况。表中指出各分区的起始地址和长度，并为每个分区设置一个标志位。当标志位为“0”时表示分区空闲，当标志位非“0”时表示分区已被占用。

#### 三、地址转换和存储保护

1、地址转换：采用静态重定位方式。

2、存储保护：处理器设置一对寄存器，即“下限寄存器”和“上限寄存器”用来存放当前进程所对应分区的下限地址和上限地址分别送入下限寄存器和上限寄存器。下限地址 $\leq$ 绝对地址 $\leq$ 上限地址，成立则执行，否则产生“地址越界”中断。09年7月单选题就考查了本知识点。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园...

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

## 第五节 可变分区存储管理（简单应用）

### 一、主存空间的分配与回收

1、分区的划分：系统初始启动时，主存储器中除操作系统占用部分外，把整个用户区看做一个大的空闲区。当有作业要装入主存储器时，根据作业对主存空间的需要量，从空闲区中划出一个与作业长度一致的分区来装入作业，剩余部分仍为空闲区。

2、主存空间的分配算法：分区分配表由两张表格组成，一张是“已分配区表”，另一张是“空闲区表”。常用的分配算法：

（1）最先适应分配算法：每次分配是总是顺序查找空闲去表，找到第一个能满足作业长度要求的空闲区，分割这个找到的空闲区，一部分分配给作业，另一部分分配给空闲区。

（2）最优适应分配算法：按作业要求从所有的空闲区中挑选一个能满足作业要求的最小空闲区，这样可保证不去分割一个更大的区域，使装入大作业时比较容易得到满足。

（3）最坏适应分配算法：这种算法总是挑选一个最大的空闲区分割一部分给作业使用，使剩下的部分不至于太小，仍可供分配使用。

3、主存空间的回收算法：归还区有下邻空闲区；归还区有上邻空闲区；归还区既有上邻空闲区又有下邻空闲区；归还区既无上邻空闲区又无下邻空闲区。

### 二、地址转换和存储保护

1、地址转换：采用动态重定位。作业执行过程中，每当取出一条指令后，就把该指令中的相对地址与基址寄存器的内容相加得到绝对地址。可变分区存储管理一般采用动态重定位的方式。为实现地址重定位和存储保护，系统设置了相应的硬件；基址/限长寄存器（或上界/下界寄存器）、加法器、比较线路等。

基地寄存器用来存放程序在内存的起始地址，限长寄存器用来存放程序的长度。处理器在执行时，用程序中的相对地址加上基址寄存器中的基地址，形成一个绝对地址，并将相对地址与限长寄存器进行计算比较，检查是否发生地址越界。

2、存储保护：基址寄存器内容 $\leq$ 绝对地址 $\leq$ 限长寄存器内容，成立则执行，否则产生“地址越界”中断。

### 三、移动技术

所谓移动是指把作业从一个存储区域移到另一个存储区域的工作。采用移动技术有两个目的：

- （1）集中分散的空闲区；
- （2）便于作业动态扩充主存。

移动可集中分散的空闲区，提高主存空间的利用率。移动也为作业动态扩充主存空间提供了方便。但是采用移动技术时必须注意下列问题：

- （1）移动会增加系统开销；
- （2）移动是有条件的。

所以，在采用移动技术的系统中，应尽可能地减少移动，以降低系统开销，提高系统效率。为此可改变作业装入主存储器的方式来达到减少移动的目的。

## 第六节 页式存储管理（简单应用）

### 一、基本原理

1、基本原理：把主存储器分成大小相等的许多区，每个区称为一块。与此对应，编制程序的相对地址也分成页，页的大小与块的大小相等。

2、相对地址构成：页号和页内地址。

### 二、存储空间分配与回收

可用一张主存分配表来记录已分配的块和尚未分配的块以及当前剩余的空闲块数。由于块的大小是固定的，所以主存分配表可简化为一张“位示图”，用 0 表示对应块为空闲，1 则为占用。

### 三、页表和地址转换

#### 1. 基本原理

页式存储管理是把主存储器分成大小相等的许多区，每个区称为一块。采用分页式存储管理时，相对地址由两部分组成：页号和页内地址。在进行存储空间分配时，总是以块为单位进行分配。

页式存储管理必须解决两个关键的问题：第一，怎样知道主存储器上哪些块已被占用，哪些块是空闲的；第二，作业信息被分散存放后如何保证作业的正确执行。

1、页表：每个作业一张页表，用来指出相对地址中页号和主存中块号的对应关系。

2、地址转换：采用动态重定位方式。每执行一条指令时，都要由地址转换机构按相对地址中的页号查页表，得到该页对应的主存块号，再按相对地址中的页内地址换算出欲访问的主存单元的绝对地址。计算办法：

(1) 方法一：绝对地址=块号\*块长+页内地址。

(2) 方法二：将块号作为绝对地址的高位，将页内地址作为绝对地址的低位，即可形成绝对地址（因为分块和分页大小是一致的）。09 年 4 月解答题终究考了本知识点。

#### 3、快表：

(1) 页表缺点：需两次访问主存。第一次按页号读出页表中对应的块号；第二次按计算出来的绝对地址访问主存。

(2) 快表：是指存放在高速缓冲存储器中的部分页表。而把存放快表的高速缓冲存储器称为相联存储器。

## 第七节 虚拟存储管理（简单应用）

### 一、什么是虚拟存储器

如果能做到只装入作业的部分信息就可以让作业开始执行，那么当主存空间小于作业需求量时，系统就可以接受该作业，进而也就可以允许逻辑地址空间大于实际的主存空间。这样就带来两个好处，第一，使主存空间能充分地利用；第二，从用户的角度来看，好像计算机系统提供了容量很大的主存储器，我们称它为虚拟存储器，简称虚存。虚拟存储器实际上是为扩大主存容量而采用的一种管理技巧。

### 二、虚拟存储器的工作原理

1、工作原理：把作业信息保留在磁盘上，当要求装入时，只将其中一部分先装入主存储器，作业执行过程中，若要访问的信息不在主存中，则再设法把这些信息装入主存。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园....

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

## 2、程序特点:

(1) 第一, 程序执行时有些部分是彼此互斥的, 即在程序的一次执行中, 执行了这部分就不会去执行另一部分

(2) 第二, 程序的执行往往具有局部性, 在一段时间里可能循环执行某些指令或多次访问某一部分的数据。

## 3、实现虚拟存储器的关键问题:

(1) 怎样知道当前哪些信息已在主存储器中, 哪些信息尚未装入主存储器中?

(2) 如果作业要访问的信息不在主存储器中, 怎样找到这些信息并把它们装到主存储器中?

(3) 在把欲访问的信息装入主存储器时, 发现主存中已无空闲块又该怎么办?

## 三、页式虚拟存储器的实现

1、实现原理: 将作业的全部信息作为副本存放在磁盘上, 作业调度选中一个作业时, 至少把作业的第一页信息装入主存储器, 在作业执行过程中欲访问不在主存储器中的页时, 再把它们装入。为此, 页表需更改, 至少应包括页号、标志、主存块号、磁盘上的位置。

## 2、页面调度:

(1) 页面调度概念: 是指采用某种算法选择一页暂时调出, 把它存放到磁盘上去, 让出主存空间, 用来存放当前要使用的页面的这一过程称为页面调度。

(2) 抖动: 也称颠簸, 是指刚被调出的页又立即要用, 因而又要把它调入; 而调入不久又被调出; 调出不久又再次被调入。如此反复, 使调度非常频繁, 以至于使大部分时间都花费在来回调度上。这种现象称为抖动。

## (3) 页面调度算法

A. 最佳调度算法 (OPT): 总是把以后不再访问的页或距当前最长时间后再访问的页先调出。是一种理想算法, 用来被用做衡量其他算法的标准, 09 年 4 月综合体就考了本知识点。

B. 先进先出调度算法 (FIFO): 总是把先进入主存储器的页面调出。

C. 最近最久未使用调度算法 (LRU): 距当前最长时间内没有使用过的页面先调出。

D. 最近最不经常使用调度算法 (LFU): 在最近一段时间内使用次数最少的页面先调出。

## 3、缺页中断率 09 年 4 月综合体就考了本知识点。

(1) 缺页中断率定义: 如果作业执行中访问页面的总次数为 A, 其中有 F 次访问的页面尚未装入主存, 故产生了 F 次缺页中断。现定义缺页中断率 f 如下:  $f = F/A$ 。

(2) 影响缺页中断率的因素: 分配给作业的主存块数; 页面大小; 程序的编制方法; 页面调度算法。

## 四、多级页表

1、原理: 建立页表时, 第一级是页面组表 (称为一级页表, 每个作业一张), 第二级是组内页面表 (称为二级页表, 每组一张)。一级页表指出二级页表的存放地址, 二级页表指出每个页在主存中的块号。

2、地址转换方法: 采用二级页表结构的系统总是把页表保存在辅助存储器中。程序执行时只需把一级页表先装入主存储器。进行地址转换时, 按相对地址中的页号 I 查一级页表, 找出对应的表项, 再根据表项中的标志位可以知道对应的二级页表是否已在主存中。若已在主存, 则可按页号 II



查二级页表中的表项,得到页所在位置(已在主存或尚未装入主存)。若二级页表尚未装入主存,则应先将其装入,再按页号 II 查找页所在位置。若页已在主存中,则根据对应的主存块号和相对地址中的页内地址得到当前要访问的主存绝对地址,否则需将该页先调入主存再进行地址转换。

3、优缺点:有利于主存空间的利用,但会增加访问主存的次数,会影响指令执行速度。

## 第八节 UNIX系统的页式虚拟存储管理(领会)

### 一、UNIX 的虚拟地址结构

1、虚拟地址:UNIX 采用页式虚拟存储管理,并把编程序时用的地址称为虚拟地址。

2、地址空间划分:系统区段、程序区段、控制区段。操作系统在系统区段运行,正文段和非共享程序都在程序区段,控制区段用来存放用户栈、核心栈、user 区等。系统区段中的程序和数据常驻主存,其余两个区段中的信息可随执行情况在主存和磁盘对换区之间换进/换出。

3、虚拟地址结构:最高两位表示访问的区段;最低 9 位表示页内地址(称为字节偏移量);中间 21 位表示页号(称为虚拟页号)。

### 二、UNIX 的页表和地址转换

1、页表结构:

(1)“V”为有效位。如果此位为 1,表示这个虚拟页已经在主存,其主存位置由物理页号指定;如果此位为 0,则进程在访问该页时硬件将产生缺页中断,由操作系统进行处理。

(2)“M”为修改位,指出该页是否被修改过。若 M 为 1,表示该页被修改过,被修改过的页面被调出时必须将其写回到磁盘上,否则不需要重写到磁盘上。

(3)“PROT”四位是访问权限指示位。进行地址转换时都要核查这四位。如果一个进程超出了规定的访问权限,硬件将产生一个“非法访问”中断。

(4)物理页号:即主存块号。

2、地址转换:

硬件为每个区段设置一对页表寄存器,分别用来存放该区段页表的起始地址和长度。在进行地址转换时,首先判断虚拟地址区段,然后找到该区段的页表寄存器,按寄存器中指示的地址就可找到该区段的页表。

只要虚拟地址在指定的长度范围内就能在页表中找到该页的登记项。检查登记项中的有效位,若为 1,则该页已在主存,把登记项中的物理页号作为绝对地址的高地址部分,把虚拟地址中的字节偏移量作为绝对地址的低地址部分。

### 三、UNIX 的页面调度

1、优化措施:

(1)一个正在与外围设备交换信息的页面或一个正在被装入的页面是不能被替换的。

(2)页面调度采用二次机会页面替换算法。

(3)为了装入一个新页而要调出一页时,要检查被调出页的修改位标志。若该页被修改过,则调出时必须把该页的内容写回磁盘上,否则就不必写回磁盘,以减少输入输出传送。

(4)系统中有一个 2 号进程,UNIX 把它称为页面守护进程,它的作用是保证有足够的空闲物理页可供使用。一般它都处于睡眠状态。

2、二次机会页面替换算法实现要点

3、页面守护进程职责

## 第五章 文件管理

### 内容概要

本章介绍了文件与文件系统的基本概念和文件管理方法，主要包括：

文件系统为用户提供按名存取的功能，用户从使用角度组织文件，用户组织的逻辑文件有两种形式：流式和记录式。

文件系统从存储介质的特性、用户的存取方式以及怎样有效地存储和检索的角度组织文件，由文件系统组织的物理文件类型可以有顺序文件、链接文件和索引文件等。

文件系统通过查找文件目录得到用户指定的文件在存储介质上的位置，读出物理文件再转换成逻辑文件传送给用户。

为了正确地实现按名存取，用户与系统必须密切配合。对用户来说，应组织好逻辑文件，为文件定义文件名、存取方式、记录格式、记录长度以及存储设备类型等；调用规定的文件操作请求使用文件。对文件系统来说，要做好存储介质初始化的工作；确定物理文件的组织形式；进行文件存储空间的分配和回收；文件目录的查找和登陆；解释执行文件操作。

考虑到系统的效率与安全，文件系统还实现记录成组与分解工作，文件保护与保密功能，精心设计文件目录的结构以及磁盘空闲块的查找方法。

在 UNIX 中，用户把文件组织成流式文件，而系统把它组织成索引文件放到磁盘上。索引表是多级索引的结构，即可快速检索短小文件，又能适应大型文件。

UNIX 采用树型目录结构，把目录项中的管理和控制信息集中索引节点中。

### 重点内容

#### （一）文件和文件系统

文件：我们把逻辑上具有完整意义的信息集合称为文件。一个文件可以是一系列的二进制数、字符或记录，它们的含义和命名都由文件的建立者定义。

文件系统对文件统一管理，目的是方便用户且保证文件的安全可靠。面向用户，文件系统主要是实现“按名存取”。为此，文件系统应由如下一些部分组成：（1）文件目录；（2）文件的组织；（3）文件存储空间的管理；（4）文件操作；（5）文件的安全措施。

#### （二）文件的存储介质

我们把可用来记录信息的磁带、应磁盘组、软磁盘片、光盘、卡片等称为存储介质，把可安装存储介质的设备称为存储设备。要把信息记录到存储介质上或从存储介质上读出信息必须启动相应的磁带机、磁盘驱动器等设备。

我们把存储介质的物理单位定义为卷，例如，一般磁带、一张软盘片、一个磁盘组都可称为一个卷。

我们把存储介质上连续信息所组成的一个区域称为块（物理记录）。块是主存储器与这些设备进行信息交换的单位。目前常用的存储设备是磁带机和磁盘机。

#### （三）文件的存取方式

从树文件信息的存取次序考虑,存取方式可以分成两种:顺序存取和随机存取。顺序存取是指对文件中的信息按顺序依次进行读写的存取方式。随机存取是指对文件中的信息不一定要按顺序读写,而是可以按任意的次序随机地读写的存取方式。

采用哪种存取方式,主要取决于两个方面的因素:

- (1) 与文件的使用方式有关;
- (2) 与存储介质的特性有关。

目前常用的存储设备是磁带机和磁盘机。

磁带机是一种适合顺序存取的存储设备。它总是从磁头的当前位置开始读写磁带上的信息。所以对存储在磁带上的文件,一般均采用顺序存取的方式。

磁盘机是一种可按指定的地址进行信息存取的设备。磁盘机上的每一块都有确定的位置和一个唯一的地址。所以磁盘机能随机读写磁盘机上任何一块的信息,具有随机存取功能。

#### (四) 文件目录

文件目录是用于检索文件的,它是文件系统实现按名存取的重要手段。文件目录由若干目录项组成。一般来说,每个目录项中应包含:有关文件存取控制的信息,有关文件结构的信息,有关文件管理的信息。

##### 1. 一级目录

一级目录是最简单的文件目录结构。这种目录结构是把一卷存储介质上的所有文件都登记在一个文件目录中。

每当建立一个新文件时就在文件目录中增加一个目录项;每当删去一个文件时就在文件目录中删去该文件的目录项。但这种结构无法解决文件重名问题。

##### 2. 二级目录

二级目录结构是为每个用户设置张用户文件目录,再用主文件目录来登记各个用户的文件目录表存放地址。

##### 3. 树形目录

大多数文件系统允许在文件目录中再建立其子目录,即形成多级目录结构。如:Windows、UNIX、MS—DOS 系统都采用多级目录结构。这种结构也称为树型目录结构,该树从根向下,每一个节点都是一个目录或一个文件。

访问文件时,必须指出文件所在的路径,路径由从根目录开始到该文件的通路上所有各级子目录名及该文件名顺序拼起来组成。通常引入当前目录的概念,将某级目录设置为当前工作目录,要访问文件时,就可从当前目录开始设置路径,称相对路径。

树形结构的优点:

- (1) 解决了重名问题
- (2) 有利于文件分类
- (3) 提高检索文件的速度
- (4) 能进行村务权限的控制

#### (五) 文件的组织

自考乐园,自考学习交流、资料共享的好去处! 自考乐园,自考人自己的家园...

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

文件的组织是指文件的构造方式。用户眼中看到的文件结构,称为文件的逻辑结构(逻辑文件)。在存储介质上文件实际的存储方式称为文件的存储结构。

### 1. 文件的逻辑结构

逻辑文件可以有两种形式:一种是流式文件,另一种是记录式文件。流式文件是指用户对文件中的信息不再划分可独立的单位,整个文件是由依次的一串信息组成。记录式文件是指用户对文件中的信息按逻辑上独立的含义再划分信息单位。每个单位称为一个逻辑记录,简称记录。如数据库文件就是一种记录式文件。

### 2. 文件的存储结构

用户的逻辑文件要存放到存储介质上时。文件系统要根据存储设备的类型、用户采用的存取方式决定文件在存储介质上的组织方式。通常文件的组织方式有三种:顺序结构、链接结构和索引结构。

#### (1) 顺序结构

一个文件被存放到连续相邻的块上,其逻辑记录顺序和物理块的顺序相一致,这类文件称顺序文件或连续文件。文件占用的每一块的物理地址及文件长(或末地址)登记在该文件目录项中。

#### (2) 链接文件

链接结构文件的逻辑记录是顺序的,但在存储空间中不必选择连续的物理块。每个磁盘块中必须留出若干单元用来存放物理块之间的链接指针。文件信息占用的第一个磁盘块的块号登记在文件目录中,链接结构中每个磁盘块中的链接指针指出了文件信息存入的下一个磁盘块的块号。

链接结构与顺序结构都只适合于顺序存取,不适宜随机访问,而下面介绍的索引结构文件适于随机访问。

#### (3) 索引结构

索引结构是实现非连续存储的另一种方法。索引结构为每个文件建立一张索引表,把指示每个逻辑记录存放位置的指针集中在索引表中。索引结构文件既可随机存取也可顺序存取,索引表的位置应登记到该文件的目录项中。

磁带上文件只能组织成顺序结构,磁盘上文件可以组织成任何一种形式。

### 3. 记录的成组和解组

把若干个逻辑记录合成一组存入一块的工作称为记录的成组。每块中的逻辑记录个数称为块因子。从一组逻辑记录中把一个逻辑记录分离出来的操作称为记录的分离。

记录的成组不仅提高了存储空间的利用率,而且还减少了启动外设的次数,提高了系统的工作效率。但成组与分解也有其缺点,记录的成组与分解是以设立主存缓冲区和操作系统增加成组与分解操作的功能为代价,来提高存储介质的利用率和减少启动设备的次数的。

## (六) 磁盘存储空间的管理

当用户要求存储文件时,就要为他分配存储空间。当用户要求删除文件时,又要收回文件占用的存储空间。如何实现存储空间的分配和回收取决于对空闲块的管理方法。通常用位示图法、空闲块表法、空闲块链法来实现对空闲块的管理。

### 1. 位示图法

用一张位示图来指示磁盘存储空间的使用情况,磁盘分块后,根据可分配的总块数决定位示图

由多少个字组成，它的每一位与一个磁盘块对应。某位为“1”状态表示相应块已被占用，为“0”状态表示所对应的块是空闲块。

## 2. 空闲块表法

系统为每个磁盘建立一张空闲块表，其中每个登记项记录一组连续空闲块的首块号和块数。这种管理方式适合采用顺序结构的文件。

## 3. 空闲块链法

把所有的磁盘空闲块用指针连接在一起构成空闲块链。分配空间时从链中取出空闲块，归还空间时把归还块加入到链中。这种管理方式不需要外加专门记录空闲块分配情况的表格。

## 考情分析

本章主要考查文件的存取方式，磁盘文件的组织，记录的成组与分解，位示图法的管理，unix的文件管理，题型一般为单选题，多选题，填空题，综合题。分值大约为15分。

例如：09年4月单选题13考查了（文件的按名存取），09年4月填空题30考查了（文件存取方式的选择），09年7月单选题13考查了（磁盘文件的组织形式），09年4月综合题40题，09年7月综合题39考查了（记录的成组与分解），09年4月简答题34考查了（位示图法管理中的公式）09年7月简答题35考查了（unix中close调用的过程）

## 串讲内容

# 第一节 文件和文件系统（识记）

## 一、文件

- 1、文件：是指逻辑上具有完整意义的信息集合。每个文件都要用一个名字作标识，称为文件名。
- 2、文件分类：分类标准不一样，文件可分为不同的类型。

## 二、文件系统的组成部分

- 1、文件系统目的：对文件统一管理，目的是方便用户且保证文件的安全可靠。面向用户，文件系统主要是实现“按名存取”。
- 2、文件系统组成部分：文件目录、文件的组织、文件存储空间的管理、文件操作、文件的安全措施。
- 3、文件操作：为了保证文件系统能正确地存储和检索文件，系统规定了在一个文件上可执行的操作，这些可执行的操作统称为文件操作。基本操作有建立文件、打开文件、读文件、写文件、关闭文件和删除文件。

# 第二节 文件的存储介质（领会）

我们把可用来记录信息的磁带、应磁盘组、软磁盘片、光盘、卡片等称为存储介质，把可安装存储介质的设备称为存储设备。要把信息记录到存储介质上或从存储介质上读出信息必须启动相应的磁带机、磁盘驱动器等设备。

我们把存储介质的物理单位定义为卷，例如，一般磁带、一张软盘片、一个磁盘组都可称为一个卷。

我们把存储介质上连续信息所组成的一个区域称为块（物理记录）。块是主存储器与这些设备进行信息交换的单位。目前常用的存储设备是磁带机和磁盘机。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园....

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部



### 第三节 文件的存取方式（领会）

#### 一、存取方式分类

从树文件信息的存取次序考虑，存取方式可以分成两种：顺序存取和随机存取。顺序存取是指对文件中的信息按顺序依次进行读写的存取方式。随机存取是指对文件中的信息不一定要按顺序读写，而是可以按任意的次序随机地读写的存取方式。

采用哪种存取方式，主要取决于两个方面的因素：09年4月解答题考了本知识点。

- (1) 与文件的使用方式有关；
- (2) 与存储介质的特性有关。

目前常用的存储设备是磁带机和磁盘机。

磁带机是一种适合顺序存取的存储设备。它总是从磁头的当前位置开始读写磁带上的信息。所以对存储在磁带上的文件，一般均采用顺序存取的方式。

磁盘机是一种可按指定的地址进行信息存取的设备。磁盘机上的每一块都有确定的位置和一个唯一的地址。所以磁盘机能随机读写磁盘机上任何一块的信息，具有随机存取功能。

#### 二、存取方式选择

1、与文件的使用方式有关：文件的性质决定了文件的使用，也就决定了存取方式的选择。如源程序文件必须按字符顺序进行存取，数据库的访问则采用随机存取方式。

2、与存储介质的特性有关：磁带机适合顺序存取的存储设备；磁盘机既可采取顺序存取方式，又可采用随机存取方式。

### 第四节 文件目录（简单应用）

#### 一、一级目录结构

- 1、基本思想：把一卷存储介质上的所有文件都登记在一个文件目录中。
- 2、要求：在文件目录中登记的各个文件都有不同的文件名。

#### 二、二级目录结构

- 1、用户文件目录：是二级目录结构中为每个用户设置的一张目录表。
- 2、主文件目录：是一张用来登记各个用户的目录表存放地址的总目录表。
- 3、优点：

(1) 采用二级目录结构后，即使不同的用户在为各自的文件命名时取了相同的名字也不会引起混乱。

(2) 采用二级目录结构可使不同的用户共享某个文件，这只要在各用户的文件目录表中使某个目录项指向共享文件存放的物理位置即可。

#### 三、树形目录结构

- 1、多级目录结构：也称树形目录结构。
- 2、绝对路径：路径名可以由从根目录开始到该文件的通路上所有各级子目录名及该文件名顺序拼起来组成，各子目录与文件名之间用“\”隔开。
- 3、相对路径：访问文件时，从当前目录开始设置路径。
- 4、树形目录结构优点：解决了重名问题；有利于文件分类；提高检索文件的速度；能进行存取

权限的控制。

#### 四、文件目录的管理

1、目录文件：通常把文件目录页作为文件保存在存储介质上，由文件目录组成的文件称为目录文件。

2、文件目录的管理：文件系统可以根据用户的要求从目录文件中找出户的当前目录，把当前目录读入主存储器。这样既不占用太多的主存空间，又可减少搜索目录的时间。

### 第五节 文件的组织（领会）

#### 一、文件的逻辑结构

1、文件的逻辑结构：用户从使用的角度来组织文件，用户把能观察到的且可以处理的信息根据使用构造文件，这种构造方式是独立于物理环境的，所以称为文件的逻辑结构。

2、逻辑文件：是指用户组织的文件，简称为文件。

3、逻辑文件分类：

（1）流式文件：是指用户对文件中的信息不再划分可独立的单位，整个文件是由依次的一串信息组成。

（2）记录式文件：是指用户对文件中的信息按逻辑上独立的含义再划分信息单位。区分概念：逻辑记录、逻辑记录号、主键、次键。

#### 二、文件的存储结构

1、文件的存储结构：文件系统从文件的存储和检索的角度来组织文件，文件系统根据存储设备的特性、文件的存取方式来决定以怎样的形式把用户文件存放到存储介质上，在存储介质上的文件构造方式称为文件的存储结构。

2、物理文件：是指存放在存储介质上的文件。

3、磁带文件的组织：

（1）组织方式：磁带机是一种顺序存取设备，因此组织在磁带上的文件都采用顺序结构。

（2）磁带文件的组成：

A. 文件头标：用来标识一个文件和说明文件的属性。

B. 文件信息：是用户逻辑文件中的信息。

C. 文件尾标：用来表示一个文件的信息结束。

→文件与文件之间用一个带标隔开，用两个带标表示磁带上的有效信息到此结束。

4、磁盘文件的组织：09年4月单选题就考了本知识点。

（1）顺序结构

优点：顺序结构适合顺序存取，其优点是存取信息的速度快、存取文件时不必每次去查找信息的存放位置，只要记住当前块号，则该文件的后继信息一定在下一块中，减少了检索时间。

缺点：磁盘存储空间利用率不高；对输出文件很难估计需多少磁盘块；影响文件的扩展。

（2）链接结构

A. 链接结构：顺序的逻辑记录被存放在不相邻的磁盘块上，再用指针把这些磁盘块按逻辑记录的顺序链接起来，便形成了文件的链接结构。

B. 链接文件：也称串联文件，是指采用链接结构的文件。

C. 注意事项：

a. 对链接文件采用顺序存取方式是高效的，采用随机存取方式是低效的。

b. 在插入或删除一个记录时，若某一块中的指针需修改，则应先把该块内容读到主存储器，再修改指针，然后把修改后的该块信息按原地址重新写回磁盘。

c. 每一个磁盘块中既要存放文件信息，又要存放用于管理的指针，这会增加文件需占用的磁盘块数。

d. 读写磁盘上的信息以块为单位。当读出一块信息后，应把其中的指针分离出来，仅把属于逻辑文件的信息传送给用户，以保证用户使用文件信息的正确性。

e. 在存取文件时，如果某个指针丢失或被破坏，则错误的指针可能指向其他文件而导致混乱。

(可采用双指针)

### (3) 索引结构

A. 索引结构：为每一个文件建立一张索引表，把指示每个逻辑记录存放位置的指针集中在索引表中。

B. 索引文件：是指采用索引结构的文件。

C. 特点：索引结构既适合顺序存取记录，又可方便地按任意次序随机存取记录，且容易实现记录的增、删和插入。但采用索引结构必须增加索引表占用的空间和读写索引表的事件。

5、存取方式与存储结构之间的关系。

## 三、记录的成组与分解

1、记录的成组：把若干个逻辑记录合成一组存入一块的工作称为记录的成组。每块中的逻辑记录个数称为块因子。

2、记录的分解：从一组逻辑记录中把一个逻辑记录分离出来的操作。

3、优缺点：

(1) 优点：提高存储空间的利用率、减少启动外设的次数、提高系统的工作效率。

(2) 缺点：是以设立主存缓冲区和操作系统增加成组与分解操作的功能为代价。

## 第六节 磁盘存储空间的管理 (领会)

### 一、位示图法

用一张位示图来指示磁盘存储空间的使用情况，磁盘分块后，根据可分配的总块数决定位示图由多少个字组成，它的每一位与一个磁盘块对应。某位为“1”状态表示相应块已被占用，为“0”状态表示所对应的块是空闲块。

确定空闲块常用的公式为

块号 = 字号 × 字长 + 位号

柱面号 = [块号 / 柱面上的块数]

磁头号 = [(块号 mod 柱面上的块数) / 盘面上的扇区数]

扇区号 = (块号 mod 柱面上的块数) mod 盘面上的扇区数

可参照第4章中页式存储空间“位示图”管理方法。

## 二、空闲块表法

系统为每个磁盘建立一张空闲块表，表中每个登记项记录一组连续空闲块的首块号和块数。空闲块数为“0”的登记项为无效登记项。

## 三、空闲块链法

把所有的磁盘空闲块用指针连接在一起构成空闲块链。分配空间时从链中取出空闲块，归还空间时把归还块加到链中。对磁盘空闲块可用单块连接法连接起来。每一个空闲块中都设置一个指向另一个空闲块的指针，最后一个空闲块中的指针为“0”。

# 第七节 基本文件操作及其使用（领会）

## 一、基本文件操作：要求掌握相应的参数和主要工作。

- 1、建立操作
- 2、打开操作
- 3、读操作
- 4、写操作
- 5、关闭操作
- 6、删除操作

## 二、文件操作的使用

- 1、读一个文件信息：打开文件\读文件\关闭文件。
- 2、写一个文件信息：建立文件\写文件\关闭文件。
- 3、删除一个文件：关闭文件\删除文件。

# 第八节 文件的安全性（领会）

## 一、文件的保护

- 1、防止天灾人祸造成的破坏：采用建立多个副本的办法。
- 2、防止系统故障造成的破坏：采用建立副本和定时转储的办法。
- 3、防止文件共享时造成的破坏：采用不允许同时使用；或允许同时使用但限制对文件使用的权限。
- 4、防止计算机病毒的侵害：采用针对各种病毒设计相应的杀毒软件，还可在二进制文件的目录中设置一般用户只能读的权限。

## 二、文件的保密

- 1、隐蔽文件目录
- 2、设置口令
- 3、使用密码

# 第九节 UNIX系统的文件管理（领会）

## 一、UNIX 的文件和文件系统

- 1、文件的逻辑结构：是由一串顺序的字符流组成的流式文件。
- 2、文件的存储结构：采用索引结构方式，即把文件按一定的长度分块后存放到磁盘，并建立索

引表。

3、文件分类：普通文件、目录文件和设备文件。

4、文件系统：分成基本文件系统和可装卸的子文件系统两部分。09 年 7 月选择题就考查了本知识点

## 二、UNIX 的文件结构

1、磁盘空间划分：分成 512 个字节的许多块。

2、文件结构：采用多级索引结构。规定每个文件的索引表使用 13 个登记项，前 10 个登记项直接指出存放文件信息的磁盘块号。如果 10 个磁盘块不够容纳该文件信息，则利用第 11 个登记项指向一个磁盘块，该磁盘块作为文件的一级间接索引，共有 128 个登记项，可分别指向 128 个磁盘块。对于大型文件还可利用第 12 和第 13 两个登记项作为二级和三级间接索引。

## 三、UNIX 的文件目录

1、目录结构：UNIX 采用树形目录结构。

2、目录文件存取权限：09 年 4 月单选题就考查了本知识点。

(1) 读：可以读该目录。

(2) 写：可以增、删这个目录中的目录项，从而改变目录内容。

(3) 执行：可以为寻找一个文件而搜索这个目录。

## 四、UNIX 的索引节点

1、索引节点概念：也称为 i 节点或 i-node，是把目录项中用于对文件进行控制和管理的信息分离出来，单独组成一个数据结构。

2、磁盘块分类：

(1) 存放索引节点的索引节点区。

(2) 存放文件信息的文件储存区。

3、索引节点区管理：对索引节点区中的磁盘块进行编号，其中每一个磁盘块可以用来存放一个索引节点。索引节点区中的第一个磁盘块存放文件系统的根目录文件的索引节点，它的索引节点号就为 1。其余的磁盘块随文件的建立和删除进行分配和回收。

4、磁盘索引节点：存放在磁盘上索引节点区中的索引节点称为磁盘索引节点。磁盘索引节点实际上是描述文件的第一个抽象数据结构，其中的 di addr 指向文件的索引表，从而可以找到文件信息的具体存储位置。

5、活动索引节点表：UNIX 在主存开辟一个索引节点缓冲区，用来建立活动索引节点表 inode[]。inode[] 有 100 个表项，能暂存 100 个索引节点。

6、目录索引：文件目录中的每个目录项由 16 个字符组成，其中 14 个字符为文件名，是该文件的外部标识，另两个字符是文件的索引节点号，它是该文件的内部标识。因此，目录索引的流程为：按文件名查找目录项→取得对应的索引节点号去找相应的磁盘索引节点→将磁盘索引节点复制到→活动索引节点表中→通过活动索引节点可找到索引表→按照索引表中的对应情况存取磁盘上的相应信息。

## 五、UNIX 的打开文件表

1、系统打开文件表 file[]：反映对被打开文件进行操作的动态信息，主要用于对文件读写操作



的控制。共有 100 个表项。

2、进程打开文件表：用来记录每个进程打开的文件情况。UNIX 系统规定每个进程最多同时可打开 15 个文件。

## 六、UNIX 的文件操作

- 1、建立文件 creat
- 2、打开文件 open
- 3、读文件 read
- 4、写文件 write
- 5、关闭文件 close 09 年 4 月解答题就考了本知识点。
- 6、系统调用 link：为一个文件再取一个新文件名
- 7、系统调用 unlink：删除文件的一个文件名。

## 七、UNIX 的文件存储空间管理（领会）

UNIX 把磁盘上的用来存放文件信息的磁盘块采用成组链接的方法进行管理。成组链接法是把空闲块分成若干组，把指向一组中各空闲块的指针集中在一起。UNIX 规定，每 100 个空闲块为一组，每组的第一个空闲块中登记下一组空闲块的磁盘物理块号和空闲块数，最后不足 100 块的那部分磁盘物理块号及块数记入专用块中。

# 第六章 设备管理

### 内容概要

本章介绍了设备管理的基本概念和主要方法。内容包括：

文件管理提供按名存取，能把用户文件保存在存储介质上，或从存储介质上取出文件信息传送给用户。要实现这一功能必须把存储介质装到相应的外围设备上并且启动外围设备工作，所以，从设备管理的角度应考虑设备的分配和启动，以配合文件管理实现信息传送。

从使用的角度，可把外围设备分成独占使用的设备和可共享使用的设备两大类。对独占使用的设备一般采用静态分配策略，根据用户指定的设备类和台数进行分配。对可共享的设备如磁盘不进行预先分配，而是根据确定的驱动调度算法来决定当前可以使用磁盘者。

启动外围设备完成输入/输出操作的过程可分为三个阶段：组织通道程序并把通道程序首地址存入通道地址字（CAW）中；用“启动 I/O”指令启动通道工作，通道执行通道程序把执行情况记录在通道状态字（CSW）中；完成输入/输出操作后形成 I/O 中断，由操作系统做出相应的处理。

为了提高独占设备的使用效率，创造多道并行工作的环境，增加单位时间内的算题量，充分利用系统的资源，操作系统借助硬件提供的中央处理器与外围设备之间的并行工作能力，采用多道程序设计技术，实现斯普林（SPOOLING）系统，通过预输入、“井管理”和缓输出为用户提供虚拟设备。

UNIX 把设备分成块设备和字符设备两大类，且把设备作为一种特别文件，每个设备都有一个文件名，并对设备就像文件那样进行操作。

本章要求了解设备管理与文件管理的合作关系，文件管理实现文件存取前的准备工作，而文件的物理存取由设备管理实现。理解怎样实现独占设备的分配和磁盘的驱动调度；怎样实现虚拟设备。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园....

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

## 重点内容

### (一) 设备管理的功能

设备管理应具有如下功能：设备的分配和回收、外围设备的启动、对磁盘的驱动调度、外部设备中断处理、虚拟设备的实现。

### (二) 外围设备的分类

从使用的角度来分析外围设备的特性，可以把外围设备分成两大类：一类是只能让一个作业独占使用的设备；另一类是可以由几个作业同时使用的设备。

### (三) 独占设备的管理

#### 1、设备的绝对号与相对号

计算机系统对每一台设备都要进行登记，且为每一台设备确定一个编号，以便区分和识别，这个确定的编号称为设备的绝对号。

由用户对自己需要使用的若干台同类设备给出的编号称为设备的相对号。

#### 2、独占设备的分配

作业申请独占设备时，应指定所需要的是什么设备。指定的方式有两种：一种是指定设备的绝对号，另一种是指定设备类、相对号。

### (四) 磁盘的驱动调度

#### 1、访问磁盘的操作时间

对磁盘的任何一个访问请求，应给出访问磁盘的存储空间地址：柱面号、磁头号、扇区号。

磁盘执行一次输入/输出操作所需要花费的时间由寻找时间、延迟时间和传送时间三部分组成。

系统往往采用一定的调度策略来决定各等待访问者的执行次序，我们把这项决定等待访问者执行次序的工作称为驱动调度，采用的调度策略称为驱动调度算法。驱动调度包括“移臂调度”和“旋转调度”两部分。

#### 2、移臂调度

根据等待访问者指定的柱面位置来决定次序的调度称为移臂调度。常用的移臂调度算法有先来先服务调度算法、最短寻找时间优先调度算法、电梯调度算法和单向扫描调度算法。

#### 3、旋转调度

根据延迟时间来决定执行次序的调度称为旋转调度。

#### 4、信息的优化分布

记录的优化分布有利于减少延迟时间，从而缩短了输入/输出操作的时间。所以，对于一些能预知处理要求的信息采用优化分布可以提高系统的效率。

### (五) 输入/输出操作的实现

现代计算机系统引入了自成独立系统的通道结构。通道的出现使计算机系统的性能得到提高，它把中央处理器（CPU）从琐碎的输入/输出操作中解脱出来，为计算机系统中各个部件能并行工作创造了条件。

正因为通道能单独地完成输入/输出操作，所以把通道也称为输入/输出处理机。完成输入/输出操作的过程是：按系统规定的通道命令格式，根据输入/输出的要求组织通道程序，并把通道程序首地址存入通道地址字（CAW）中；利用硬件提供的“启动 I/O”指令启动通道工作，启动成功后由通

更多**优质**自考资料尽在百度贴吧**自考乐园**俱乐部

(<http://tieba.baidu.com/club/5346389>) 欢迎♥加入...欢迎♥交流...止不住的惊喜等着你.....

道执行通道程序,按通道命令的要求完成一次输入/输出操作,并把通道程序的执行情况记录在通道状态字(CSW)中,通道完成一次输入/输出操作后形成I/O中断,由操作系统作出相应处理。

### (六) 缓冲技术

在操作系统中,把利用缓冲区来缓解处理器与外围设备之间工作速度不匹配的矛盾而采用的技术称为缓冲技术。操作系统中采用的缓冲技术有单缓冲技术、双缓冲技术和缓冲池技术。

### (七) 虚拟设备

实现虚拟设备必须有一定的硬件和软件为基础。对硬件来说,必须配置大容量的磁盘,要有中断装置和通道,具有中央处理器与通道并行工作的能力。对操作系统来说,应采用多道程序设计技术。

操作系统中实现虚拟设备的功能模块是在计算机控制下通过联机的外围设备同时操作来实现其功能的,这个功能模块称为斯普林系统。英文缩写为SPOOLING。

更多**优质**自考资料尽在百度贴吧**自考乐园**俱乐部

(<http://tieba.baidu.com/club/5346389>) 欢迎♥加入...欢迎♥交流...止不住的惊喜等着你.....

### (八) UNIX 的设备管理

UNIX把设备分成块设备和字符设备两大类,且把设备作为一种特别文件。每个设备都有一个文件名,并对设备就像文件那样进行操作。

UNIX中的输入/输出操作是通过缓冲区来传送信息的。在采用的磁盘缓冲技术中,利用缓冲区保存已从磁盘上读出的信息或欲写入磁盘的信息。这样,当再次需要使用这些信息时就不必启动磁盘而可直接从缓冲区中得到,减少了输入/输出操作所花费的时间。

### 考情分析

本章主要考查设备的相对号和绝对号,访问磁盘的操作方法,移臂调度的方法,信息的优化分布,通道的结构,unix的设备管理,有单选题、填空题,分值大约为10分。

例如09年填空题31题考查了(独占设备的特点),09年7月填空题30题考查了(磁盘访问过程),09年4月填空题24题考查了(移臂调度的几种方法)。09年7月单选题16考查了(通道程序的结构),

### 串讲内容

## 第一节 设备管理的功能(识记)

#### 1、外围设备分类:(从功能上分)

(1) 存储型设备:磁盘机、磁带机等。

(2) 输入/输出型设备:显示器、输入机、打印机等。

#### 2、输入/输出操作:是指主存储器与外围设备之间的信息传送操作。

(1) 对存储型设备:输入/输出操作的信息传送单位为“块”。

(2) 对输入/输出型设备:输出/输出操作的信息传送单位为“字符”。

#### 3、设备管理的主要功能:

(1) 实现对外围设备的分配与回收。

(2) 实现外围设备的启动。

自考乐园,自考学习交流、资料共享的好去处!自考乐园,自考人自己的家园...

俱乐部id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部id,可以直接进入俱乐部

- (3) 实现对磁盘的驱动调度。
- (4) 处理外围设备的中断事件。
- (5) 实现虚拟设备。

## 第二节 外围设备的分类 (领会)

- 1、独占设备：是指在作业执行期间只允许一个作业独占使用的设备。如输入机、打印机等。
  - 2、可共享设备：是指可让若干个作业同时使用的设备。如磁盘机等。(同时使用：是指一个作业尚未撤离，另一个作业即可使用，但每一时刻仍只有一个作业能启动磁盘，允许它们交替地启动)
- 以上是从使用的角度来分。

## 第三节 独占设备的管理 (领会)

### 一、设备的绝对号与相对号

- 1、设备的绝对号：为了对设备进行管理，计算机系统对每一台设备都要进行登记，且为每一台设备确定一个编号以便区分和识别，这个确定的编号称为设备的绝对号。
- 2、设备的相对号：由用户对自己需要使用的若干台同类设备给出的编号称为设备的相对号。

### 二、独占设备的分配

- 1、设备申请的指定方式：
  - (1) 指定设备的绝对号：系统就必须把对应的设备分配给作业，且不可替代。
  - (2) 指定设备类和相对号：可以实现用户编制程序时使用的设备与实际能占用的设备无关。
- 2、设备的独立性：是指用户程序中使用由“设备类、相对号”定义的逻辑设备，可以实现用户编制程序时使用的设备与实际能占用的设备无关，该种使用设备的特性称为设备的独立性。
- 3、设备的独立性优点：
  - (1) 系统只要从指定的那一类设备中找出“好的且尚未分配的”设备来进行分配。
  - (2) 万一用户使用的设备出了故障，系统就可以从同类设备中找另一台“好的且尚未分配的”设备来替换。
- 4、独占设备的分配：系统可设置设备分配表，指出系统所配置的独占设备类型、数量以及分配和使用的情况等。该表由两部分组成：
  - (1) 设备类表：每一类独占设备在设备类表中占一个登记项。表项包括设备类、拥有的总台数、现存台数、设备表始址。
  - (2) 设备表：每一台设备在设备表中占一个登记项。表项包括绝对号、好/坏、已/未分配、占用作业名、相对号。

## 第四节 磁盘的驱动调度 (综合应用)

### 一、访问磁盘的操作时间

- 1、输入/输出操作所需花费的时间：
  - (1) 寻找时间：磁头在移动臂带动下移动到指定柱面所需的时间。
  - (2) 延迟时间：指定扇区旋转到磁头位置所需的时间。
  - (3) 传送时间：由指定的磁头把磁道上的信息读到主存储器或把主存储器中信息写到磁道上所

需的时间。由于块的长度一样，因此传送时间是固定的。

## 2、驱动调度：

(1) 移臂调度：是指根据等待访问者指定的柱面位置来决定次序的调度。目标是尽可能地减少寻找时间。

(2) 旋转调度：是指根据延迟时间来决定执行次序的调度。目标是尽可能地减少延迟时间。

## 二、移臂调度

- 1、先来先服务调度算法
- 2、最短寻找时间优先调度算法
- 3、电梯调度算法
- 4、单向扫描调度算法

## 三、旋转调度

1、若干请求者要访问同一磁头下的不同扇区：旋转调度总是对先到达读写磁头位置下的扇区进行信息传送。

2、若干请求者要访问不同磁头下的不同编号的扇区：旋转调度总是对先到达读写磁头位置下的扇区进行信息传送。

3、若干请求者要访问不同磁头下具有相同编号的扇区：旋转调度根据磁头号可从中任意选择一个磁头进行读/写操作，其余的请求者必须等磁盘再次把扇区旋转到磁头位置时才有可能被选中。

## 四、信息的优化分布：可减少驱动调度。

# 第五节 输入/输出操作的实现（识记）

## 一、通道结构和通道程序

1、通道结构：也称输入/输出处理机，是计算机系统中用来连接外围设备，并能独立工作的系统。

2、通道的连接：一个中央处理器可以连接多个通道；一个通道可以连接多个设备控制器；一个设备控制器可以连接同类型的多台设备。

3、通道的基本工作模式：当有输入输出请求时，中央处理器先执行“启动 I/O”指令，启动指定通道上的指定设备。当启动成功，通道按规定的要求通过设备控制器控制外围设备进行操作。这时中央处理器就可执行其它任务，并与通道并行工作，直到输入输出操作完成，由通道发出操作结束的“I/O 中断”时中央处理器才暂停当前的工作，转去处理 I/O 中断事件。

4、通道命令 CCW：每一条通道命令规定了设备的一种操作。操作系统可以用若干条通道命令来规定通道执行一次输入/输出操作应做的工作。若干条通道命令组成了一个通道程序。

5、通道地址字 CAW：用来存放通道程序首地址的主存固定单元。通道系统可以依据该地址去取相应的通道命令。

6、通道状态字 CSW：用来存放通道和设备执行情况的主存固定单元。

(1) 通道命令地址：用来指示下一条通道命令在主存中的地址。（因为通道地址字只存放通道程序的首地址）

(2) 设备状态：用来记录设备控制器和设备的工作情况。



(3) 通道状态：用来记录通道的工作情况。

(4) 剩余字节个数：是指最后一次执行的那条通道命令在操作结束后还剩余多少字节未传输完。

## 二、外围设备的启动

1、输入/输出操作的过程：

(1) 准备阶段

(2) 启动 I/O 阶段

(3) 结束处理阶段

2、设备处理的一致性：在具有通道结构的计算机系统中，从启动外围设备到完成输入/输出操作都没有考虑不同类型的物理设备的特性，都采用了统一的方法在进行处理。这种不考虑设备的具体物理特性（实际上设备特性已隐含在通道程序中）的处理方法称为设备处理的一致性。

## 三、I/O 中断事件的处理

1、作用：I/O 中断是通道和中央处理器协调工作的一种手段。

2、I/O 中断事件分类：

(1) 操作正常结束：

(2) 操作异常结束：

A. 设备故障

B. 设备特殊

# 第六节 缓冲技术（领会）

## 一、单缓冲技术

1、缓冲技术：是指操作系统中，把利用缓冲区来缓解处理器与外围设备之间工作速度不匹配的矛盾而采用的技术。

2、单缓冲技术：是指操作系统在主存储器的系统区中只设立一个缓冲区。

## 二、双缓冲技术

双缓冲技术是利用两个缓冲区来完成输入输出操作的工作。两个缓冲区交替使用。

## 三、缓冲池技术

操作系统可以在主存中设置一组缓冲区，把这一组缓冲区称为缓冲池。缓冲池中的各缓冲区是系统的公共资源，可供各进程共享，并由操作系统统一分配和管理。

# 第七节 虚拟设备（领会）

## 一、为什么要提供虚拟设备

1、虚拟设备概念：操作系统利用共享设备来模拟独占设备的工作，当系统只有一台输入设备和一台输出设备的情况下，可允许两个以上的作业并行执行，并且让每个作业都感觉到获得了供自己独占使用的输入设备和输出设备。我们说，操作系统采用的这种技术为用户提供了“虚拟设备”。

2、为什么要提供虚拟设备：我们已经知道像输入机、打印机等独占使用的设备采用静态分配方式，既不能充分利用设备，又不利于提高系统效率。主要表现为：

(1) 占有输入机和打印机的作业，只有一部分时间在使用它们，在其余时间这些设备处于空闲状态。在设备空闲时不允许其他作业去使用它们，因此不能有效地利用这些设备。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园...

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id, 可以直接进入俱乐部

(2) 当系统只配有一台输入机和一台打印机时,就不能接受两个以上要求使用输入机和打印机的作业同时执行,不利于多道并行工作。

(3) 这些独占设备大多是低速设备,在作业执行中往往由于等待这些设备的信息传输而延长了作业的执行时间。

## 二、虚拟设备的实现

1、基本条件:实现虚拟设备必须有一定的硬件和软件条件为基础。对硬件来说,必须配置大量的磁盘,要有中断装置和通道,具有中央处理器与通道并行工作的能力。对操作系统来说,应采用多道程序设计技术。

2、实现原理:把一批作业的全部信息通过输入设备预先传送到磁盘上等待处理。在多道程序设计系统中,可从磁盘上选择若干个作业同时装入主存储器,并让它们同时执行。由于作业的信息已全部在磁盘上,故作业执行时不必再启动输入机读信息,而可以从共享的磁盘上读取各自的信息。把作业产生的结果也存放在磁盘上,而不直接启动打印机输出。直到一个作业得到全部结果而执行结束时,才把该作业的结果从打印机输出。

3、实现技术:

(1) 输入井和输出井:

A. 输入井:存放作业的初始信息。

B. 输出井:存放作业的执行结果。

(2) 斯普林系统:操作系统中实现虚拟设备的功能模块是在计算机控制下通过联机的外围设备同时操作来实现其功能的,这个功能模块的英文缩写为 SPOOLING,所以称它为斯普林系统。

A. 预输入程序:任务是把作业流中每个作业的初始信息传送到输入井保存以备作业执行时使用。

B. 井管理程序:井管理程序又分成井管理读程序和井管理写程序两个功能。当作业请求从输入机上读文件信息时,就把任务转交给井管理读程序。由它从输入井读出信息供作业执行时使用。当作业请求从打印机上输出结果时,就把任务转交给井管理写程序。由它把产生的结果保存到输出井中。

C. 缓输出程序:缓输出程序负责查看输出井中是否有待输出的结果信息。若有,则启动打印机把结果文件打印输出。

(3) 数据结构

A. 作业表:用来登记输入井中的各个作业的作业名、作业状态、作业拥有的文件数,以及预输入表和缓输出表的位置等。

B. 预输入表:每个作业有一张预输入表,用来登记该作业的初始信息中的各个文件,指出各文件的文件名、传输文件信息时使用的设备类型、文件的长度以及文件的存放位置等。

C. 缓输出表:对每个作业设置一张缓输出表,用来登记该作业产生的结果文件。作业产生的结果也按链接结构组织成文件存放在输出井中。

(4) 输入井中作业的状态:

A. 输入状态:预输入程序启动了输入机正在把该作业的信息传输到输入井。

B. 收容状态:该作业的信息已经存放在输入井中,但尚未被选中执行。

C. 执行状态:作业已被选中并装入主存储器开始执行。

D. 完成状态：作业已执行结束，其执行结果在输出井中等待打印输出。

## 第八节 UNIX 的设备管理（领会）

### 一、UNIX 的设备和设备文件

1、块设备和字符设备：按设备与主存之间信息交换的物理单位来分。

(1) 块设备：也称存储设备，以块为单位与主存交换信息，是用来存储信息的设备。如磁盘机、磁带机等。

(2) 字符设备：也称输入/输出设备，以字节为单位与主存交换信息，是用来接收外部信息或把处理好的信息传向外部的设备。如终端、打印机等。

2、主设备号和次设备号：

(1) 主设备号：是 UNIX 对每一类设备给出的一个编号。

(2) 次设备号：是为了标识某一具体设备，UNIX 对每一台设备给出的一个编号。请求设备工作时，必须给出主设备号和次设备号。系统根据主设备号决定由哪个驱动程序工作，驱动程序再根据设备号确定控制哪台具体的设备工作。

3、设备文件：

UNIX 把设备当作特别文件。每个设备特别文件也都有一个索引节点。索引节点中的文件类型被定义为“块”或“字符”，一则区别于其他文件，二则用以确定该索引节点代表的是块设备文件还是字符设备文件。设备文件登记在根目录下的 dev 子目录中。把设备定义成文件后，就可以对设备象文件那样操作，对设备的使用也是通过系统调用 open, close, read, write 来进行。这些系统调用的内部实现与普通文件的系统调用的实现相似，但要增加一些与设备有关的处理过程。

### 二、UNIX 的块设备缓冲技术

1、UNIX 采用的缓冲技术类型：系统设置一批缓冲区，构成系统缓冲区池。对写操作，驱动程序分配一个缓冲区，先把信息从用户空间复制到缓冲区，再启动设备把缓冲区中信息保存起来。对读操作，启动设备后把指定的信息传送到一个缓冲区中，再将信息复制到指定的用户空间。这样做的目的是为了减少启动设备的次数。

2、缓冲区构成：

(1) 缓冲区数据区：用于存放文件信息。

(2) 缓冲区控制块：用于缓冲区的管理。主要管理信息：

A. 状态标志 b flags：描述该缓冲区当前的状态。

B. 队列指针：共有四个队列指针。具体包括：av\_forw：空闲缓冲区队列前向指针；av\_back：空闲缓冲区队列后向指针；b\_forw：设备缓冲区队列前向指针；b\_back：设备缓冲区队列后向指针。

C. 设备号 dev：缓冲区数据区中信息所属设备的设备号，包括主设备号和次设备号。

D. 字节数 b bcount：请求传送的字节数。

E. 块号 b blkno：指定设备上的块号。

F. 存放地址 b addr：存放信息的主存起始地址。

3、缓冲区队列：

(1) 空闲缓冲区队列：空闲缓冲区队列又称空闲 av 链。它是由指针 av\_forw 和 av\_back 形成

双向链接队列。该队列中所有缓冲区的状态标志 BUSY 位为 0，表示都是空闲缓冲区。系统初始启动时将所有缓冲区按序号挂在 av 链上。

(2) 设备缓冲区队列：设备缓冲区队列又称设备 b 链。它是由指针 b\_forw 和 b\_back 形成的双向链接队列。该队列链接设备正在使用或已经使用过的缓冲区。可为每个设备设立一个设备缓冲区队列。

#### 4、缓冲区队列的管理原则：

(1) 当需要一个缓冲区时，总是从空闲缓冲区队列的队首取一个缓冲区。一个被使用过的缓冲区释放时，总是排入空闲缓冲区队列的队尾。

(2) 一个缓冲区被分配用于读写某磁盘块信息时，这个缓冲区就从空闲缓冲区队列退出，并链入该设备的设备缓冲区队列，该缓冲区状态标志中应置上 BUSY（表示缓冲区正在“忙”）标志。

(3) 当缓冲区中的信息传送到用户空间后，或当用户信息写到了缓冲区且对缓冲区置了 DELWR（延迟写标志，表示缓冲区中内容尚未写到磁盘上）标志后，这样的缓冲区可以释放。此时便清除缓冲区中的 BUSY 标志，把它链到空闲缓冲区队尾，同时仍把它保留在该设备的设备缓冲区队列中。这里首先要注意的是，虽然缓冲区被链入空闲缓冲区队列，但缓冲区中的信息仍是有效的。

(4) 分配缓冲区时，总是摘取空闲缓冲区队列中的第一个缓冲区。但如果该缓冲区有 DELWR 标志（延迟写标志），则表示该缓冲区信息尚未写回磁盘，暂不能分配。必须将它从空闲缓冲区队列退出，且提出 I/O 请求，由设备驱动程序把该缓冲区内容写到相应设备的指定磁盘块上。待输入输出操作完成后，清除该缓冲区的 DELWR 标志且将它重新链入空闲缓冲区队列的队尾，同时仍把它保留在原设备缓冲区队列中，其目的是当需要时仍然不必启动设备而可直接使用该块中的信息。

(5) 如果一个缓冲区既在空闲缓冲区队列中，又在设备缓冲区队列中，而该缓冲区被分配移作它用，则要让它从空闲缓冲区队列和原设备缓冲区队列退出，且链入新的设备缓冲区队列。

#### 5、缓冲区的检索：由设备管理中的缓冲区检索模块 getblk 完成。

(1) 根据设备号找到该设备的设备缓冲区队列。如果能找到一个缓冲区，其缓冲区控制块中的 b\_blkno 与文件系统提供的磁盘块号相同，则说明该磁盘块所需的缓冲区已经存在，因此不必再分配缓冲区。但是要根据该缓冲区是否正在忙而区别对待。

(2) 如果在设备缓冲区队列中找不到相应的缓冲区，那么就从空闲缓冲区队列中分配一个缓冲区。

## 第七章 进程同步与进程通信

### 考情分析

本章主要考查进程顺序执行的特点，与时间有关的错误，用 PV 操作实现进程的同步于互斥，进程的通信，线程的概念，题型一般为单选题，多选题，填空题，综合题，有时概念会在简答题中考查。分值大约为 10 分。

例如：09 年 4 月单选题 19 题考查了（进程的交互性），09 年 7 月填空题 31 题考查了（与时间有关错误的概念），09 年 4 月综合题 41 题考查了（用 PV 操作实现进程的同步于互斥）09 年 4 月简答题 36 题考查了（线程的概念）

### 串讲内容



## 内容概要

本章介绍了进程同步和通信的基本概念和方法。主要包括：

在顺序执行指令的处理器上，进程的执行是按顺序的，即进程具有顺序性。但是，在计算机系统中往往有若干个进程请求执行，当一个进程的执行没有结束前允许其他进程也开始执行，则说这些进程是可同时执行的。可同时执行的进程交替地占用处理器，若系统中存在一组可同时执行的进程，则说该组进程具有并发性，把这些进程称为“并发进程”。

并发进程相互间可以是无关的，也可以是有交互的。有交互的并发进程一定共享某些资源，由于并发进程执行的相对速度受自身和外界的因素影响，也受进程调度策略的限制，因此并发进程在访问共享资源时可能会出现与时间有关的错误。并发进程在访问共享资源时可以有两种关系，一种是竞争关系，另一种是协作关系。当并发进程存在竞争关系时，必须互斥地使用共享资源；当并发进程存在协作关系时，必须互通消息。把这两种关系分别称为“进程的互斥”和“进程的同步”，进程间正确的互斥和同步能避免出现与时间有关的错误。

把并发进程中与共享变量有关的程序段称为“临界区”。与某共享变量有关的每个进程都有各自的临界区，但这些临界区都涉及共同的变量，故把它们称为是“相关临界区”。进程的互斥是指并发进程互斥地进入相关临界区，即每次只允许一个进程进入自己的临界区，当有一个进程在它的临界区执行时就不允许其他任何进程进入各自的与其相关的临界区，直到该进程退出临界区为止。进程的同步是指进程使用共享资源时必须互通消息，即只有接到了指定的消息后进程才能去使用共享资源，如果进程没有得到指定消息，虽然此时无进程在使用共享资源，该进程仍不能去使用共享资源，直到消息到达为止。实际上，进程互斥是进程同步的特例，故经常把实现进程同步和进程互斥的机制统称为“同步机制”。

PV 操作是一种简便的同步机制，它包括两个不可中断的过程，分别称为“P 操作原语”和“V 操作原语”。PV 操作是对信号量实施操作，若把信号量与共享资源对应起来，则用 PV 操作可以实现进程的同步和进程的互斥。从教材的实例中可以总结信号量的物理含义（假定信号量是 S 表示）如下：

$S > 0$  时，S 表示可使用的资源数或表示可使用资源的进程数；

$S = 0$  时，S 表示无资源可供使用或表示不允许进程再进入临界区；

$S < 0$  时， $|S|$  表示等待使用资源的进程个数或表示等待进入临界区的进程个数。

于是，根据 PV 操作的性质，任何进程在使用共享资源前应调用 P 操作。当有可以使用的资源或允许使用资源时 ( $S > 0$ )，调用 P (S) 后不会成为等待，进程可以使用资源。当无资源可使用或不允许使用资源时 ( $S \leq 0$ )，调用 P (S) 后进程必然等待。任何进程可调用 V 操作来归还共享资源使用权，当  $S > 0$  时调用 V (S) 后使可用的资源数加 1 或使可用资源的进程数加 1。当  $S \leq 0$  时调用 V (S) 后将释放一个等待使用资源者或释放一个等待进入临界区者实现进程互斥时，用一个信号量与一组相关临界区对应，这些进程在同一个信号量上调用 P 操作和 V 操作来实现互斥。实现进程同步时，每一个消息与一个信号量对应，进程在不同信号量上调用 P 操作以测试自己需要的消息是否到达，在不同信号量上调用 V 操作可把不同的消息发送出去。



除 PV 操作同步机制外, 其他的同步机制还有“管程”(monitor) 和“忙时等待”等, 有兴趣的话可参阅其他书籍。

进程可用信件来传送大量信息, 一个进程可以发送一封信, 把信息告诉其他进程或请求其他进程协助工作。发送信件者必须在信件中指出信件送给谁(或哪个信箱)以及所要的信息(或信息存放地址), 这样才能使收信者收到信件和取道信息。信件的传送要求依赖专门的通信机制, 通信机制由一些通信原语组成, 其中 send 原语和 receive 原语是最基本的通信原语。

多线程技术是操作系统的发展趋势, 它能提高计算机系统的性能。

## 重点内容

### (一) 进程的顺序性和并发性

#### 1. 进程的顺序性

目前使用的计算机的基本特点是处理器顺序执行指令。进程的顺序性就是指进程在顺序处理器上的执行是严格按序的。

当一个进程独占处理器顺序执行时, 具有封闭性和可再现性两个特性。

封闭性是指进程执行的结果只取决于进程本身, 不受外界影响。可再现性是指当进程重复执行时, 必定获得相同的结果。

#### 2. 进程的并发性

可同时执行的进程是指这些进程执行在时间上是重叠的, 即一个进程的工作没有全部完成之前, 另一个进程就可以开始工作。

若系统中存在一组可同时执行的进程, 则该组进程具有并发性, 可同时执行的进程称为并发进程。

#### 3. 并发进程间的关系

并发进程间的关系可以是无关的, 也可以是有交互的。并发进程间无关是指它们是各自独立的, 即如果一个进程的执行不影响其他进程的执行, 且与其他进程的进展情况无关; 则并发进程间有交互是指一个进程的执行可能影响其他进程的执行结果, 即一个进程的执行依赖其他进程的进展情况。有交互的并发进程一定共享某些资源。

### (二) 与时间有关的错误

有交互的并发进程可能会同时使用共享资源, 如果对这种情况不加控制, 在使用共享资源时会出错。典型的例子是由观察者和报告者两个进程管理交通路口的自动计数系统。观察者识别卡车并计数, 报告者定时把观察者的计数值打印输出, 然后把计数值清 0。当这两个进程并发执行时, 若报告者进程执行时无卡车通过, 则能正确统计某时间段内的卡车数量; 若报告者进程执行时有卡车通过, 统计结果会产生错误。

### (三) 临界区与 PV 操作

#### 1. 临界区

临界区是指并发进程中与共享变量有关的程序段。

#### 2. 相关临界区

相关临界区是指并发进程中涉及相同变量的那些临界区。

只要若干并发进程的相关临界区互斥执行, 就不会造成与时间有关的错误。

### 3、管理相关临界区的三点要求

- (1) 一次最多让一个进程在临界区执行;
- (2) 任何一个进入临界区执行的进程必须在有限的时间内退出临界区;
- (3) 不能强迫一个进程无限地等待进入它的临界区。

### 4、PV 操作

P 操作和 V 操作是两个在信号量上进行操作的过程, PV 操作就是由这两个操作组成的。

假设用 S 表示信号量, 用 P(S) 和 V(S) 表示 P、V 操作。

(1) P(S) 的过程如下:

S 值减 1;

若  $S < 0$ , 则将该进程排入等待 S 的队列;

若  $S \geq 0$ , 则当前进程继续执行。

(2) V(S) 的过程如下:

S 值加 1;

若  $S \leq 0$ , 从等待 S 的队列中移出一个进程并加入到就绪队列, 当前进程继续执行;

若  $S > 0$ , 则当前进程继续进行。

不可中断的过程称为原语, 所以 P 操作和 V 操作也称为 P 操作原语和 V 操作原语。

### 5. 用 PV 操作管理临界区

用一个信号与一组涉及共享变量的相关临界区相对应。信号量的初值定为 1, 任何一个进程进入临界区前, 先调用 P 操作, 退出临界区时调用 V 操作, 就能满足管理相关临界区的三个要求。

## (四) 进程的互斥

进程的互斥是指当有若干进程都要使用某一共享资源时, 任何时刻最多只允许一个进程去使用该资源, 其他要使用它的进程必须等待, 直到该资源的占用者释放了该资源。

实际上共享资源的互斥使用就是使得并发进程互斥地进入临界区。

## 第一节 进程的顺序性与并发性 (领会)

### 一、进程的顺序性

1、进程的顺序性: 是指进程在顺序处理器上的执行是严格按序的, 即按照程序规定的操作顺序, 只有在前一个操作结束后才能开始后继操作。

2、进程顺序执行的特性:

(1) 封闭性: 进程执行的结果只取决于进程本身, 不受外界影响。也就是说, 进程执行的结果与其执行的速度无关。

(2) 可再现性: 进程重复执行时, 必定获得同样的结果。

### 二、进程的并发性

1、并发性: 在一个进程的工作没有全部完成之前, 另一个进程就可以开始工作, 我们说这些进程是可同时执行的, 或称它们具有并发性, 并且把可同时执行的进程称为并发进程。

2、并发进程间的关系:

(1) 无关: 如果一个进程的执行不影响其它进程的执行结果, 也不依赖其它进程的进展情况,

即它们是各自独立的，则说这些进程相互之间是无关的。

(2) 有交互：如果一个进程的执行要依赖其它进程的进展情况，或者可能会影响其它进程的执行结果，则说这些进程相互之间是有交互的。有交互的进程并发执行时，执行结果与其执行的相对速度有关。例如，如果输入进程尚未把一批数据全部读入，处理进程就对其进行加工的话，那么其结果就会出错。因而，进程的并发执行会破坏“封闭性”和“可再现性”。

## 第二节 与时间有关的错误(简单应用)

造成进程执行不正确的因素是与进程占用处理器的时间、执行的速度以及外界的影响有关。这些因素都与时间有关，所以把它们称为与时间有关的错误。09年7月解答题就考了本知识点

## 第三节 临界区与PV操作(领会)

### 一、临界区

1、临界区：是指并发进程中与共享变量有关的程序段。

2、相关临界区：是指并发进程中涉及相同变量的那些临界区。

说明：如果有进程在临界区执行时，不让另一个进程进入相关的临界区执行，就不会形成多个进程对相同的共享变量交叉访问，于是就可避免出现与时间有关的错误。

3、相关临界区的管理要求：

(1) 一次最多一个进程能够进入临界区。当有进程在临界区执行时，其它想进入临界区执行的进程必须等待。

(2) 不能让一个进程无限制地在临界区执行，即任何一个进入临界区的进程必须在有限的时间内退出临界区。

(3) 不能强迫一个进程无限制地等待进入它的临界区，即有进程退出临界区时应让一个等待进入临界区的进程进入它的临界区执行。

### 二、PV操作(是不可中断的过程，即原语)

1、P操作(占用资源的过程)：将信号量S减去1，若结果小于0，则把调用P(S)的进程置成等待信号量S的状态。

2、V操作(释放资源的过程)：将信号量S加1，若结果不大于0，则释放一个等待信号量S的进程。

### 三、用PV操作管理临界区

...P(S);

{临界区Ci};

V(S);

...

## 第四节 进程的互斥(综合应用)

### 一、用PV操作实现进程的互斥

进程的互斥：是指当有若干进程都要使用某一共享资源时，任何时刻最多只允许一个进程去使用该资源，其它要使用它的进程必须等待，直到该资源的占用者释放了该资源。

## 二、读者/写者问题

1、共享文件：在计算机系统中，把可供多个进程使用的文件称为共享文件。我们把想读文件信息的进程称为读者（读进程），把想修改文件内容的进程称为写者（写进程）。

2、共享文件的管理方式：

（1）不允许多个进程同时使用共享文件：这种管理方式限制了每次只有一个进程可以使用文件。实际上这仍是一个互斥使用共享资源的问题。

（2）允许多个进程同时使用共享文件。要注意使用规则。

## 第五节 进程的同步(综合应用)

### 一、协作

1、进程 A 把一个记录存入缓冲区后，应向进程 B 发送“缓冲区中有等待处理的记录”的消息。

2、进程 B 从缓冲区取出一个记录后，应向进程 A 发送“缓冲区中的记录已取走”的消息。

3、进程 A 只有在得到进程 B 发送来的“缓冲区中的记录已取走”的消息后，才能把下一个记录存入缓冲区，否则进程 A 等待，直到消息到达。

4、进程 B 只有在得到进程 A 发送来的“缓冲区中有等待处理的记录”的消息后，才能从缓冲区中取出记录并加工，否则进程 B 等待，直到消息到达。

### 二、用 PV 操作实现进程的同步，07 年 4 月综合体就考了本知识点

1、进程的同步：是指并发进程之间存在一种制约关系，一个进程的执行依赖另一个进程的消息，当一个进程没有得到另一个进程的消息时应等待，直到消息到达才被唤醒。

2、同步机制：要实现进程的同步就必须提供一种机制，这种机制应能把其它进程所需的消息发送出去，也能测试自己需要的消息是否到达。把能实现进程同步的机制称为同步机制。

3、同步与互斥概括：进程的同步与进程的互斥都涉及到并发进程访问共享资源的问题。从进程互斥和进程同步的讨论中，我们看到，进程的互斥实际上是进程同步的一种特殊情况。实现进程互斥时用 P 操作测试是否可以使用共享资源，这相当于测试“资源可使用”的消息是否到达；用 V 操作归还共享资源，这相当于发送了“共享资源已空闲”的消息。因此互斥使用资源的进程之间实际上也存在一个进程等待另一个进程发送消息的制约关系。所以，经常把进程的互斥与进程的同步统称为进程的同步，把用来解决进程互斥与进程同步的机制（如 PV 操作）统称为同步机制。

4、同步与互斥的混合问题：若一组涉及共享资源的并发进程执行时不仅要等待指定的消息到达，而且还必须考虑资源的互斥使用，那么，就既要实现进程的同步，又要实现进程的互斥。这样一类问题就是同步与互斥的混合问题。

## 第六节 进程通信(领会)

### 一、通信机制

1、进程通信：是指通过专门的通信机制实现进程间交换大量信息的通信方式。

2、通信机制：采用高级通信方式时，进程间用信件来交换信息。一个正在执行的进程可以在任何时刻向其它进程发送信件，一个正在执行的进程也可以在任何时刻向其它进程索取信件。

3、信件构成：

（1）发送者名：为发送信件进程的进程名。

自考乐园, 自考学习交流、资料共享的好去处! 自考乐园, 自考人自己的家园...

俱乐部 id: **5346389** (请牢记它哦~在百度贴吧的搜索框中输入俱乐部 id,可以直接进入俱乐部

(2) 信息（或信息存放的地址和长度）：指要传送给某一进程的信息，若信息量大，则可将信息存放在某个缓冲区中，在信件中指出缓冲区的起始地址和信息长度。

(3) 等/不等回信：表示信件发送者是否等信件接收者的回信。

(4) 回信存放地址：若需要等回信，应指明回信的存放地址。

#### 4、通信原语：

(1) 发送原语（send）：应给出两个参数，一个是信件或者信件存放地址；另一个是信送到哪里。

(2) 接收原语（receive）：应给出两个参数，一个是从哪里取信；另一个是取出的信存放在哪里。

#### 5、通信方式：

(1) 直接通信方式：固定在一对进程之间进行的通信。

(2) 间接通信方式：以信箱为媒体来进行的通信。

## 二、间接通信

#### 1、信箱构成：

(1) 信箱说明：包括可存信件数、已有信件数、可存信件的指针等参数。

(2) 信箱体：存放信件的区域。

#### 2、间接通信的通信规则：

(1) 若发送信件时信箱已满，则应把发送信件的进程置成“等信箱”状态，直到信箱有空时才被释放。

(2) 若取信件时信箱中无信，则应把接收信件的进程置成“等信件”状态，直到信箱中有信件时才被释放。

#### 3、间接通信中通信原语的功能及实现要求：

(1) send (N, M)

A. 功能：把信件 M 送到指定的信箱 N 中。

B. 实现要求：查指定信箱 N。若信箱未满，则按“可存信件的指针”把信件 M 存入信箱且释放“等信件”者；若信箱已满，则把发送信件进程置为“等信箱”状态。

(2) receive (N, Z)

A. 功能：从指定信箱 N 中取出一封信，存到指定的地址 Z 中。

B. 实现要求：查指定信箱 N。若信箱中有信，则取出一封信存于 Z 中，且释放“等信箱”者；若信箱中无信，则把接收信件进程置为“等信件”状态。

## 第七节 UNIX 中的进程同步与进程通信(领会)

### 一、UNIX 中的进程同步

#### 1、用 wait 和 exit 实现同步：

UNIX 的进程在用户态执行时，可使用系统调用 fork 创建新进程，被创建的进程与创建者构成了父子进程的关系。因为父进程创建子进程的目的是要子进程去完成一些特定的任务，所以父进程要等待子进程的执行，于是用系统调用 wait 让自己等待。等待期间，子进程按父进程的要求执行任



务,直到任务完成时,用系统调用 `exit` 来终止自己,且释放父进程。父进程被释放后,继续执行后继操作。所以,当进程在用户态执行用户程序时,系统调用 `wait` 和 `exit` 是实现同步的主要手段。

## 2、用 `sleep` 和 `wakeup` 实现同步:

UNIX 中,在用户态执行的进程请求操作系统为其服务时,就转换到核心态执行操作系统程序。进程在核心态执行时会出现各种等待事件。无论哪种原因引起了等待事件,都要使用系统调用 `sleep` 让进程进入睡眠状态。当等待结束时,使用系统调用 `wakeup` 来唤醒进程,使被唤醒的进程结束等待,而成为就绪状态。所以,当进程在核心态执行系统程序时,由系统调用 `sleep` 和 `wakeup` 实现同步。

## 二、UNIX 中的进程通信

### 1、管道机制:允许进程间按先进先出的方法传送消息。

(1) 无名管道 `pipo`: 适用于一个用户有同一祖先的父子进程间的通信。

所谓无名管道实际上是连接进程间的可共享文件,称为 `pipe` 文件。一个进程可以通过也称为 `pipe` 的系统调用来建立一个无名管道,然后其子进程可与该进程共享该管道,对 `pipe` 文件进行读写操作。一个进程可以把信息写入 `pipe` 文件,而另一个进程可从 `pipe` 文件中读取信息,但这种读写操作需由管道机制加上一些读写同步等来控制。`shell` 中用符号 “|” 表示对 `pipe` 文件的操作,故管道机制的一般形式是: `P1 | P2`,该操作的左分量(即进程 `P1`)的输出信息是管道 `pipe` 的输入,而管道 `pipe` 的输出则是该操作的右分量(即进程 `P2`)的输入信息。

(2) 命名管道 `FIFO`: 适用于不同用户的进程间的通信。

命名管道实际上是一个冠有文件名的管道文件。用户既可以使用 `shell` 命令中的 `mknod` 命令来产生一个有名管道文件,也可以在程序中使用系统调用 `mknod` 来创建一个有名管道文件。命名管道的使用方式与无名管道的使用方式不同。对命名管道的使用就象对普通文件的使用一样,要通过文件操作来使用。首先必须建立,读写之前先打开,通信结束后要关闭。利用有名管道文件进行通信时,通信的发送者用“只写”方式打开,通信的接收者用“只读”方式打开。对被打开的有名管道文件,进程可按打开的方式对该文件进行读或写。在读写的过程中管道机制要对读写操作进行同步控制,以保证信息传输的正确性。

## 2、消息缓冲机制:

(1) 消息缓冲机制: UNIX 中的消息缓冲机制是利用缓冲区来传输消息的。由系统统一管理一组缓冲区,其中每一个缓冲区都可用来存放一个消息。当一个进程要发送消息时,首先向系统申请一个缓冲区;然后再把组织好的消息存入缓冲区;再把存有消息的缓冲区链接到消息队列中。所有这些工作可通过调用消息缓冲机制所提供的系统调用来完成。接收消息的进程也可通过系统调用从消息队列中取用消息,从缓冲区取出消息后,就应释放该缓冲区。

(2) 缓冲区的管理信息: 发送消息的进程名、消息长度、需传输的消息正文、下一个消息缓冲区指针。

## 第八节 线程概念(领会)

### 一、什么是线程

1、线程概念: 是进程中可独立执行的子任务。

## 2、线程属性:

(1) 每个线程有一个唯一的标识符和一张线程描述表。线程描述表记录了线程执行时的寄存器和栈等现场状态。

(2) 不同的线程可以执行相同的程序,即同一个服务程序被不同的用户调用时,操作系统为它们创建成不同的线程。

(3) 同一进程中的各个线程共享分配给进程的主存地址空间。

(4) 线程是处理器的独立调度单位,多个线程是可以并发执行的。

(5) 一个线程被创建后,便开始了它的生命期,直至终止。线程在生命期内会经历等待态、就绪态和运行态等各种状态变化。

3、线程和进程的区别:进程是资源分配单位,而线程是调度和执行单位。每个进程都有自己的主存空间,同一个进程中的各线程共享该进程的主存空间,进程中的所有线程对进程的整个主存空间都有存取权限。

## 二、进程与线程

### 1、进程缺点:

(1) 每个进程要占用一个进程控制块和一个私有的主存区域,开销较大;

(2) 进程之间的通信必须由通信机制来完成,速度较慢;

(3) 进程增多会给调度和控制带来复杂性,增加了死锁的机会。

### 2、线程优点:

(1) 创建线程无须另外分配资源,因而创建线程的速度比创建进程的速度快,且系统开销小。

(2) 线程间的通信在同一地址空间中进行,故不需要额外的通信机制,使通信更简便,信息传递速度也更快。

(3) 线程能独立执行,能充分利用和发挥处理器与外围设备并行工作的能力。

# 第八章 死 锁

## 内容概要

本章主要介绍了死锁的基本概念以及对解决死锁的几种方法。主要包括:

若系统中存在一组进程,它们中的每一个进程都占用了某种资源而又都在等待其中另一个进程所占用的资源,这种等待永远不能结束,则说系统出现了“死锁”。

死锁是由于进程竞争互斥使用资源且又互不相让形成的,所以,系统中形成死锁一定同时保持了4个必要条件,即互斥使用资源、占用并等待资源、不可抢夺资源和循环等待资源。但要注意这是死锁的必要条件,而不是充分条件。

解决死锁问题可以有3种方式:

(1) 死锁的防止。预先确定一些资源分配策略,进程按规定申请资源,系统按预定的策略进行分配,这些功能策略均能使死锁的4个必要条件中的一个条件不成立,从而使系统不发生死锁。

(2) 死锁的避免。当进程提出资源申请时,系统动态测试资源分配情况,仅当能确保系统安全时才把资源分配给进程,银行家算法虽然很保守,但它可保证系统时时刻刻都处于安全状态。

(3) 死锁的检测。多资源的申请和分配不加限制,只要有剩余的资源就可把资源分配给申请者,

这样就可能出现死锁，于是系统要定时运行一个“死锁检测程序”，若检测时没有发现死锁则系统可继续工作；若检测时发现死锁则必须先解除死锁再继续工作。

在一个实际的操作系统中要兼顾资源的使用效率和安全可靠，对不同的资源可以采用不同的分配策略，往往采用死锁的防止、避免和检测的混合策略，以使整个系统能处于安全状态不出现死锁。

### 考情分析

本章主要考查死锁的必要条件，资源分配图，死锁的防止，死锁的避免中的银行家算法，题型一般为单选题，多选题，填空题，简答题。分值大约为 5 分。

例如：09 年 4 月单选题 20 考查了（资源的安全分配），09 年 7 月单选题 19 题考查了（死锁的防治），09 年 7 月单选题 20 题考查了（资源的安全分配），09 年 7 月填空题 32 考查了（死锁的相关问题）。

### 串讲内容

## 第一节 死锁的形成(领会)

1、死锁概念：若系统中存在一组进程（两个或多个），它们中每个进程都占用了某种资源，又都在等待已被该组进程中的其它进程占用的资源，如果这种等待永远不能结束，则说系统出现了死锁，或者说这组进程处于死锁状态。

2、死锁的形成：

- (1) 与进程对资源的需求有关；
- (2) 与进程的执行速度有关；
- (3) 与资源的分配策略有关。

## 第二节 死锁的特征(领会)

一、死锁的必要条件：（即死锁产生一定会成立的条件）

- 1、互斥地使用资源：每个资源每次只能给一个进程使用。
- 2、占有且等待资源：一个进程占有了某些资源后又申请新资源而得不到满足时，处于等待资源的状态，且不释放已占资源。
- 3、不可抢夺资源：任何一个进程不能抢夺另一个进程所占的资源。
- 4、循环等待资源：相互等待已被其它进程占用的资源。

必要条件说明：以上四个条件仅仅是必要条件而不是充分条件，即只要发生死锁，则这四个条件一定同时成立，如果其中的一个或几个条件不成立，则一定没有死锁。但反之不然，即若这四个条件同时成立，系统未必就有死锁存在。

二、资源分配图：（是判断死锁产生与否的方法）

## 第三节 死锁的防止(简单应用)

一、互斥条件

要使互斥使用资源的条件不成立，唯一的办法是允许进程共享资源。但部分硬件设备的物理特性是改变不了的，所以要想破坏“互斥使用资源”这个条件经常是行不通的。

## 二、占有并等待条件

1、静态分配资源：是指进程必须在开始执行前就申请自己所要的全部资源，仅当系统能满足进程的全部资源申请要求且把资源分配给进程后，该进程才开始执行。

2、释放已占资源：仅当进程没有占用资源时，才允许它去申请资源。因此，如果进程已经占用了某些资源而又要再申请资源，那么按此策略的要求，它应先归还所占的资源，归还后才允许申请新资源。（如进程要占用磁带机，又要申请打印机，可以先启动磁带机工作，完成后先释放磁带机再申请打印机。由于申请者是在归还资源后才申请新资源，故不会出现占有了部分资源再等待其它资源的现象。）

## 三、不可抢夺条件

为了使这个条件不成立，我们可以约定如下：如果一个进程已占有了某些资源又要申请新资源 R，而 R 已被另一进程 P 占用因而必须等待时，则系统可以抢夺进程 P 已占用的资源 R。具体做法如下：

（1）若进程 A 申请的资源 R 尚未被占用，则系统可把资源 R 分配给进程 A。

（2）若进程 A 申请的资源 R 已被进程 B 占用，则查看进程 B 的状态。如果进程 B 处于等待另一个资源的状态，那么就抢夺进程 B 已占的资源 R，并把 R 分配给进程 A，适当的时候再把 R 归还给进程 B 使用；否则让进程 A 处于等待资源 R 的状态。

（3）一个等待资源的进程只有在得到自己所申请的新资源和所有被其它进程抢夺去的老资源后，才能继续执行。

## 四、循环等待条件

对资源采用按序分配的策略可使循环等待资源的条件不成立。按序分配资源是指对系统中所有资源排一个顺序，对每一个资源给出一个确定的编号，规定任何一个进程申请两个以上资源时，总是先申请编号小的资源，再申请编号大的资源。

# 第四节 死锁的避免(简单应用)

## 一、安全状态

1、安全状态概念：如果操作系统能保证所有的进程在有限的时间内得到需要的全部资源，则称系统处于安全状态，否则说系统是不安全的。显然，处于安全状态的系统不会发生死锁，而处于不安全状态的系统可能会发生死锁。

2、资源分配算法：在分配资源时，只要系统能保持处于安全状态，就可避免死锁的发生。故每当有进程提出分配资源的请求时，系统应分析各进程已占资源数、尚需资源数和系统中可以分配的剩余资源数，确定是否处于安全状态。如果分配后系统仍然能维持安全状态，则可为该进程分配资源，否则就暂不为申请者分配资源，直到其它进程归还资源后再考虑它的分配问题。

## 二、银行家算法（07 年综合体考查了本知识点）

- 1、当一个用户对资金的最大需求量不超过银行家现有的资金时，就可接纳该用户；
- 2、用户可以分期贷款，但贷款总数不能超过最大需求量；
- 3、当银行家现有的资金不能满足用户的尚需贷款数时，可以推迟支付，但总能使用户在有限的时间里得到贷款；

4、当用户得到所需的全部资金后，一定能在有限时间里归还所有的资金。

## 第五节 死锁的检测(领会)

### 一、死锁的检测方法

#### 1、每类资源中只有一个资源：

如果每类资源中只有一个资源，则可以设置两张表格来记录进程使用和等待资源的情况。一张为占用表，记录进程占用资源的情况。

另一张为等待表，记录进程正在等待资源的情况。任一进程申请资源时，若该资源空闲，则把该资源分配给申请者，且在占用表中登记，否则把申请者登入等待表中。死锁检测程序定时地检测这两张表。如果发现有循环等待资源的进程，则表明有死锁出现。

#### 2、资源类中含有若干个资源：

(1) 初始检测：找出资源已经满足的进程，即不再申请资源的进程。若有这样的进程，则它们一定能在有限的时间内执行结束，且归还所占的资源。所以可以把它们所占的资源与系统中还剩余的资源加在一起作为可分配的资源，同时对这些进程置上标志。

(2) 循环检测：检测所有无标志的进程，找出一个尚需资源量不超过系统中的可分配资源量的进程。若能找到，则只要把资源分配给该进程，就一定能在有限时间内收回它所占的全部资源。故可把该进程已占的资源添加到可分配的资源中，同时为该进程置上一个标志。重复执行第二步，直到所有进程均有标志，或无标志的进程尚需资源量均超过可分配的资源量。

(3) 结束检测：若进程均有标志，表示当前不存在永远等待资源的进程，也即系统不处于死锁状态。若存在无标志的进程，表示系统当前已有死锁形成，这些无标志的进程就是一组处于死锁状态的进程。检测结束，解除检测时所设置的所有标志。

### 二、死锁的解除

#### 1、终止进程

(1) 终止涉及死锁的所有进程

(2) 一次终止一个进程

#### 2、抢夺资源

从涉及死锁的一个或几个进程中抢夺资源，把夺得的资源再分配给卷入死锁的其他进程，直到死锁解除。应注意三个问题：

(1) 抢夺哪些进程的哪些资源

(2) 被抢夺者的恢复

(3) 进程的“饿死”