



自动化学院 2022 级

大学 计算机

知识点汇编

学业指导部出品

内部资料，仅供参考

二〇二二年十一月

第1章 计算与计算思维

1.1.1 略

1.1.2 现代计算机

1. ENIAC：第一台电子数字计算机。

2. 计算机的发展

(1) 第一代电子计算机：

△采用电子管代替机械齿轮或电磁继电器作开关元件。

△采用二进制代替十进制。

△程序可以储存。

△输入输出装置主要用穿孔卡。

(2) 第二代电子计算机：

△晶体管作为主要电子元件。

△采用磁芯存储器作主存，磁盘和磁带做辅存。

△出现了许多现代计算机体系结构的特性。

△编程语言得到快速发展。

△进一步扩大计算机的应用范围。

(3) 第三代电子计算机：

△采用集成电路。

△半导体存储器淘汰了磁芯存储器。

△普遍采用了微程序设计技术。

△系统软件与应用软件都有了很大的发展。

△针对不同的计算机系统，解决软件兼容问题。

(4) 第四代电子计算机：

△微处理器或超大规模集成电路取代了中小规模集成电路。

△从计算机系统结构来看，第四代计算机是第三代的扩展与延伸。

1.1.3 计算及技术应用

1. 计算科学：

(1) 数值计算——详见书 p10

(2) 数字图像处理——详见书 p10-p11

(3) 网络计算——详见书 p11

2. 计算技术的发展：

(1) 并行计算——详见书 p12

(2) 云计算——详见书 p13

(3) 边缘计算——详见书 p13-p14

1.1.4 未来计算机

1. 计算机多样化：

量子计算机——详见书 p15-p16

DNA 计算机

生物计算机

2. 计算机微型化——详见书 p16

1.2 计算机思维

计算机思维是运用计算学科的基础概念进行问题求解、系统设计以及人类行为理解的过程，它涵盖计算机广度的一系列思维活动。计算机思维、理论思维和

实验思维称为三大思维。

1.2.1 科学思维

也叫科学逻辑，它是人类实践活动的产物，是真理在认识的统一过程中，对各种科学的思维方法有机整合。其必须遵循三个基本原则：

1. 严密的逻辑性原则。
2. 方法论原则。
3. 历史性原则。

1.2.2 计算机思维的定义

1. 计算思维的六个特征：

- △计算机思维是概念化思维，不是程序化思维。
- △计算机思维是综合的基础技能，而不是机械技能。
- △计算思维是人的思维不是计算机的思维。
- △计算思维是思想，不是人造品。
- △计算思维是数据和工程互补融合的思维，不是单纯的数学性思维。
- △计算思维是面向所有人、所有专业领域的思维。

2. 计算机维在计算机学科中的体现：

- △“0 和 1”的思维。
- △“指令和程序”的思维。
- △“递归”的思维。
- △计算系统的进化思维。
- △学科融合的思维。
- △网络化的思维。

1.3 计算技术在中国的发展

1.3.2 中国的计算机——详见书 p26-p30

1.3.3 中国计算机技术的发展——详见书 p30-p33

1. 操作系统领域。
2. 计算机智能软件。
3. 机器学习。
4. 大数据知识工程。
5. 数据中心网络。
6. 区块链技术。

第 2 章 计算机理论基础

2.1 数制

2.1.1 进位计数制

1. 数制的特点

数值包括技术符号和进位规则：

- (1) 数码：表示基本数制大小的不同数字符号。
- (2) 基数：在该进位制中用到的数码个数。
- (3) 位权：每位的值对应该位上的数码乘以固定的数，即这位的位权。
- (4) 进位制：多位数码中每一位的构成以及从低位到高位进位的规则成为进位计数制，简称进位制。

2. 常用的数制

- (1) 十进制：按照逢十进一的规则进行计数，其特点：

△采用 0、1、2、3、4、5、6、7、8、9 共十个数码

△任何十进制 N 可以表示为一个按权展开多项式

$$N = \pm \sum_{i=-m}^{n-1} a_i 10^i$$

(2) 二进制：按照逢二进一的规则进行计数，其特点：

△采用 0、1 两个数码

△可以按权展开的多项式将二进制数转换成对应的十进制数

$$N = \pm \sum_{i=-m}^{n-1} a_i 2^i$$

(3) 八进制：按照逢八进一的规则进行计数，其特点：

△采用 0、1、2、3、4、5、6、7 数码

△按权展开多项式可转为十进制

$$N = \pm \sum_{i=-m}^{n-1} a_i 8^i$$

(4) 十六进制：按照逢十六进一的规则进行计数，其特点：

△采用 0、1、2、3、4、5、6、7、8、9、A、B、C、D、E、F 十六个数码

△按权展开多项式可转为十进制

$$N = \pm \sum_{i=-m}^{n-1} 16^i a_i$$

书写时为了区别数制：

(1) 在数字右下角注明数制

(2) 在数字后面加字母来区别：

加字母 B 表示为二进制数，

加字母 O 表示为八进制，

加字母 D 或不加字母表示为十进制数，

加字母 H 表示为十六进制数。

3. 数制的转换

(1) 十进制转二进制

十进制整数转换二进制整数：除 2 逆序取余。

十进制小数转换二进制小数：乘 2 顺序取整。

(2) 二进制与八进制转换

从小数点开始分别把整数和小数部分按每三位一组进行划分，不足三位，用 0 补足。然后用一位八进制数代换三位二进制数。

(3) 八进制转换为二进制数

一位八进制数可直接转换为 3 位二进制数，每位二进制数用三位二进制数来代替。

(4) 二进制与十六进制转换

从小数点开始分别把整数和小数部分按每四位一组进行划分, 不足四位, 用 0 补足。然后用一位十六进制数代换四位二进制数。

(5) 十六进制转换为二进制数

一位十六进制数可直接转换为 4 位二进制数, 每位二进制数用四位二进制数来代替。

2.1.2 计算机中的数制

1. 二进制的算数运算

(1) 加法运算规则:

$0+0=0$; $0+1=1$; $1+0=1$; $1+1=10$ (产生进位)

(2) 减法运算规则:

$0-0=0$; $1-0=1$; $1-1=0$; $0-1=1$ (产生借位)

(3) 乘法运算规则:

$0 \times 0=0$; $0 \times 1=0$; $1 \times 0=0$; $1 \times 1=1$

(4) 除法运算规则:

$0 \div 0=0$; $0 \div 1=0$; $1 \div 1=1$; $1 \div 0=0$ (无意义)

2. 二进制的逻辑运算

(1) 与运算, 规则如下:

$0 \wedge 0=0$ $0 \wedge 1=0$ $1 \wedge 0=0$ $1 \wedge 1=1$

(2) 或运算, 规则如下:

$0 \vee 0=0$ $0 \vee 1=1$ $1 \vee 0=1$ $1 \vee 1=1$

(3) 非运算, 规则如下:

非 $0=1$ 非 $1=0$

(4) 异或运算, 规则如下:

$0 \oplus 0=0$ $0 \oplus 1=1$ $1 \oplus 0=1$ $1 \oplus 1=0$

2.2 数据储存的组织方式

2.2.1 数据的单位

1. 位(bit): 计算机存储数据的最小单元, 它表示二进制的 1 位。

2. 字节(Byte): 是计算机用于描述存储和传输的单位, 每 8 个位组成 1 个字节, 即 $1B=8b$ 。

3. 字和字长: 以一个整体来处理或运算的二进制数称为一个计算机字, 一个字通常分为若干个字节, 而这组二进制数位数就是字长。

2.2.2 数据的存储

1. 储蓄单元: 一般以 8 个二进制作为一个存储单元, 也就是一个字节, 又称为字节存储单元。

2. 存储容量

常用的字节计数单位:

$1\text{Byte}=8\text{bit}$

$1\text{KB}=1024\text{Byte}$

$1\text{MB}=1024\text{KB}$

$1\text{GB}=1024\text{MB}$

$1\text{TB}=1024\text{GB}$

2.2.3 存储编址 详见书 p48

2.3 数据表示

2.3.1 数值数据

数值型数据由数字、小数点、正负号和表示乘幂的字母 E 组成。

1. 符号数值化

(1) 无符号数是指表示数值大小，不考虑符号的数据，它在计算机中的数值表示采用二进制，且把所有二进制位都解释为数值。

(2) 带符号数在计算机中用一位二进制的 0 或 1 来区别数值的正负，通常符号放在二进制的最高位，称为符号位，以 0 代表符号“+”，以 1 代表符号“-”，这种采用二进制表示形式的把符号位数值化的数称机器数，而符号没有数值化的实际二进制数称真值。

2. 小数点处理 详细图文见书 p50-p52

(1) 定点数：小数点位置固定的数称为定点数，定点数又分为定点整数（纯整数）和定点小数（纯小数）。假设机器数按 1 字节（8 位二进制位）进行表示，定点小数将小数点位置固定在数值部分的最左端，但容易出现溢出，丢失精度等问题。

(2) 浮点数：小数点位置不固定的数称为浮点数，浮点数既有整数部分又有小数部分，其中小数点在数中的位置是浮动。浮点数的最大的特点是比定点数表示的数值范围要大，可解决数据溢出，丢失精度等问题。

3. 带符号数机器数的表示

(1) 原码

一种简单的机器数表示法，它的最高位是符号位，正数的符号位用 0 表示，负数的符号位用 1 表示，其余位有效数值是二进制数的绝对值，其特点：

△表示直观易懂，与真值转换容易

△0 有两种不同的表示形式，给使用带来了不便

△加减运算复杂

(2) 反码

整数的反码与其原码相同，负数的反码是对原码除符号位外的各位取反（即 0 变 1，1 变 0），符号位保持为 1。

(3) 补码

正数的补码与其原码相同，负数的补码是除符号位外的数值部分按位取反后，再在末位加 1。

2.3.2 字符数据

(1) BCD 码：即用二进制位表示十进制，最常见的是 BCD 编码称为 8421 码，该编码方式中从左到右 4 位二进制代表的权值分别固定为 8、4、2、1。（将数字用 84291 码表示时需要把每一位数字都用四位二进制表示）

(2) ASCII 码：使用 7 位二进制数编码来表示所有的英文大写和小写字母，数字从 0 到 9、标点符号，以及在美式英语中使用的特殊控制符，共可表示 128 个可打印字符及控制字符。

(3) UCS/Unicode 标准：将全世界现代书面文字所使用的所有字符和符号集中在同一个字符集中进行了统一的编码。

(4) 汉字编码：

汉字国标码：国家标准 GB2312-80 提供了中华人民共和国国家标准信息交换用汉字编码，简称国标码，该字符集的内容由三部分组成：

△各类符号、各类数字以及各种字母、包括英文、俄文、罗马字母、日文平假名与片假名、拼音符号和制表字符。

△常用汉字

△二级常用汉字

汉字字形码：汉字存储在计算机内采用机内码，但输出时必须转换成字形码，再根据字形码输出汉字。字形码又称汉字字模，用于在显示器或打印机上输出各种文字和符号，有两种记录方法：点阵码和矢量码。

汉字输入码：也称外码，根据汉字字形各异、含义不同的特点，利用数字、符号或字母完成的表形、表义、表音的汉字输入编码方案，可大致分为：

△数字编码

△拼音类输入法

△拼形类输入法

△音形结合类输入法

汉字机内码：又称“汉字 ASCII 码”指计算机内部存储、处理加工和传输汉字时所用的由 0 和 1 符号组成的代码。

注：汉字国标码=汉字区位码+2020H

汉字机内码=汉字国标码+8080H

2.3.3 声音数据

1.声音数字化 详见书 p57

(1) 主要参数有：

采样频率：声音数字化的过程中，每秒采集的标本个数。

1) 量化精度：声音数字化过程中，每个声音样本转换成的二进制位个数。

2) 声道数：在存储、传送或播放时相互独立的声音数目。

3) 单位数据量：声音单位数据量也称为码率，每秒声音数据率由下式表示：

数据率（码率）=采样频率×量化位数×声道数

2. 声音的压缩：压缩率是指数据被压缩之前的大小和压缩之前的大小之比。

3. 声音的符号化 详见书 p59

2.3.4 图像数据

图像根据记录方式的不同可分为模拟图像和数字图像两大类：

1.模拟图像 详见书 p59

2.数字图像

数字图像的主要参数包括：

(1) 图像的分辨率：图像的像素是按照行的列排列的，像素的列数称为水平分辨率、行数称为垂直分辨率。图像的分辨率是由“水平分辨率×垂直分辨率”来表示的。

(2) 图像的像素深度

(3) 图像的数据量

如果图像未经压缩处理，一幅图像的数据量为：

图像数据量=水平分辨率×垂直分辨率×像素深度

(4) 图像压缩

2.3.5 视频数据 详见书 p62-p63

1. 帧

2. 镜头

3. 数据量

4. 视频的压缩

2.4 数据结构

2.4.1 逻辑结构

1. 集合：集合结构是一种松散的逻辑结构，集合中的任何两个数据元素之间无逻辑关系。
2. 线性结构：是最常用最简单的数据结构，是一个有序数据元素的集合，其特点是数据元素之间存在着“一对一”的线性关系，在一个线性结构中的数据元素类型相同。
3. 树形结构：是一种非线性数据结构，所有数据元素之间有明显层次特性和分支结构。
4. 图形结构：是一种复杂的非线性结构，数据元素之间存在着多对多的关系，即每个元素可以有零个或多个直接前趋，零个或多个直接后继。

2.4.2 存储结构 详见书 p65-p66

1. 顺序存储
2. 链式存储
3. 索引存储结构

2.4.3 基本操作 详见书 p66

2.4.4 典型的数据结构

1. 线性表：是最基本、最简单、也是最常用的一种数据结构。线性表属于线性结构，采用顺序存储和链式存储两种方式。 详见书 p67

- (1) 顺序表
- (2) 线性链表

2. 栈：是限定仅在表的一端进行插入或删除操作的线性表，按照先进后出的原则存储数据。 详见书 p67-p68

- (1) 顺序栈
- (2) 链栈

3. 队列：是一种特殊的线性表，只允许在表的前端进行删除操作，而在表的后端进行插入操作。 详见书 p68-p69

- (1) 顺序列
- (2) 链队列

4. 二叉树：详见计算专题

第3章 计算机技术

3.1 计算机体系及结构

3.1.1 图灵理论模型 详见书 p77

1. 阿兰·麦席森·图灵是英国著名数学家和逻辑学家，他对电子计算机的主要贡献有：

- (1) 建立图灵理论模型
- (2) 提出定义机器智能的图灵测试

3.1.2 冯·诺依曼结构

1. 冯·诺依曼计算机的基本特点：

△计算机由输入设备、输出设备、运算器、存储器和控制器五大部件组成

(1) 控制器：控制器是整个计算机的指挥中心。它负责协调输入输出设备的操作和内存访问，完成针对指令的分析、判断，发出控制信号，使计算机各部分自动、连续、协调动作，确保系统作为一个整体进行正确运行。

(2) 运算器：运算器是计算机对信息进行加工处理的核心部件之一。它在控

制器的控制下与内存交换信息，负责进行各类基于二进制数的基本算术运算、逻辑运算、比较、移位、逻辑判断等各种操作。

(3) 存储器：存储器是计算机的记忆装置，用于存放原始数据、中间数据、最终结果和处理程序。为了对存储的信息进行管理，计算机中的各种信息都要存储在存储器中。

(4) 输入设备：输入设备是向计算机输入信息的装置，用于把原始数据和处理这些数据的程序输入计算机系统中。不论信息的原始形态如何，输入计算机中的信息都使用二进制来表示。

(5) 输出设备：输出设备的主要任务是将计算机工作过程中处理的中间或最终结果信息，以用户熟悉、方便的形式输出。

△数据和指令均采用二进制形式表示

△采用存储程序方式

△计算机的工作原理 详见书 p79

3.2 计算机主要部件

3.2.1 中央处理器（CPU）

中央处理器主要由控制单元（控制器）、算术逻辑单元（运算器）以及寄存器为核心组成，是读取、解释计算机指令并执行指令处理计算机软件中的数据。

CPU 主要性能指标：

- 1.主频：即 CPU 工作的时钟频率，主频越高，CPU 的工作越快。
- 2.外频：是主板上提供的一个基准节拍供各个部件使用，主板提供的节拍称为外频。
- 3.倍频：CPU 主频相对外频的倍数。
- 4.数据总线宽度：决定了 CPU 与内存、输入/输出设备之间一次数据传输的信息量。
- 5.地址总线宽度：即地址总线位数，决定 CPU 访问的存储容量。
- 6.多核：单个裸片上具有多个可见的处理器，并且这些处理器各自拥有独立的控制和工作状态，互相之间无需共享关键资源

3.2.2 存储器

1.存储器构成：存储器是计算机系统的记忆储备，是计算机信息存储的核心，是计算机不可少的部件之一。它主要存放二进制数“0”和“1”形式的程序和数据。

2.存储器分类：

按存取速度分：高速、中速、低速存储器。

按存储材料分：半导体、磁记录、激光存储器。

按功能分：寄存器、高速缓冲存储器、主存储器、外存储器、后备存储器等。

按存储方式分：随机存储器、顺序存储器、只读存储器等。

3.存储器结构层次

(1) 寄存器存储器：

寄存器是 CPU 内部的元件，包括通用寄存器、专用寄存器和控制寄存器，是有限存储容量的高速存储部件，是存储层次的最顶端。

特点：拥有非常高的读写速度，可与 CPU 匹配，但存储容量很小。

用处：主要用来存放地址、数据及运算的中间结果。

(2) 高速缓冲存储器（Cache）：

高速缓冲存储器是位于内存和 CPU 之间高速小容量存储器，用于解决 CPU 与主

存之间速度不匹配问题。

特点：运算速度高于内存储器，增加 Cache 可以提高 CPU 的读写速度。

用处：主要存放 CPU 近期要执行的指令和数据。配合适当的调度算法可提高系统的处理速度，从而解决高速 CPU 与速度相对较慢的主存的矛盾。

注意：增加 Cache 不会改变内存存储器的容量。

(3) 主存储器：

CPU 能够直接访问的存储器，用以存放当前运行的程序和数据，根据基本功能分为只读存储器 (ROM) 和随机存储器 (RAM)。

随机存储器 (RAM)

△存储器工作时，存储在其中的数据可反复利用，只有向存储器写入新数据时存储器中内容才被更新。

△信息随着计算机的断电而自动消失，因此 RAM 是计算机处理数据的临时存储区。

(4) 只读存储器 (ROM)

△它只能读出事先所存数据固态半导体存储器，一旦储存资料就无法再将其改变或删除，最大特点是在电源中断后信息也不会消失或受到破坏。

(5) 外存储器：

又称辅助存储器，存放当前 CPU 暂时不用的程序和数据，属于外部设备。

①硬盘 详见书 p84

主要性能指标：

△容量

△转速

△平均寻道时间

②光盘存储器 详见书 p84

③U 盘：全名闪存驱动器 (USB flash disk，简称 U 盘) 采用 Flash 存储器 (闪存) 技术通过 USB 接口与计算机连接，实现即插即用

△U 盘特点：小巧便于携带、存储容量大，安全可靠性好，使用寿命长，无任何机械式装置，抗震性能极强。还具有防潮防磁、耐高低温等特性。

④固态硬盘：简称 SSD，又称固盘，是用固态电子存储芯片阵列而制成的硬盘，在便携式计算机中代替传统硬盘。虽然成本较高但在逐渐普及。

固态硬盘的特点：低功耗、无噪音、抗震动、低热量、体积小、工作温度范围大。但由于固态电子存储芯片都有一定的写入寿命，在寿命到期后数据会读不出来且难以修复。

与传统硬盘的相同与不同：

技术不同：固态硬盘利用 NAND 存储器，再配合适当的控制芯片就可以制造

外形相同。

尺寸：尺寸上固态硬盘有 2.5 英寸和 3.5 英寸，而硬盘主要为 3.5 英寸。

接口：接口上固态硬盘普遍采用 SATA-3.0 接口、MSATA 接口、PCI-E 接口、M.2 接口；而硬盘驱动器的接口主要为 SATA、mSATA 或 NGFF 等。

3.2.3 外部设备

1. 输入设备：

(1) 键盘

(2) 指点设备：①鼠标 ②轨迹球 ③触摸板

(3) 扫描仪 扫描仪的主要技术指标包括分辨率、色彩位数、扫描速度、扫描速度、扫描幅面、接口

(4) 其他输入设备

2.输出设备:

(1) 显示器 显示器的主要技术指标包括尺寸、分辨率
显示器必须配备显卡才能构成完整的显示系统,同时显卡还有图像处理能力,可协助 CPU 工作,提高整体的运行速度。

显卡分为:

①集成显卡:优点是功耗低、发热量小,但性能相对略低,且固定在主板或 CPU 上

②独立显卡;优点是单独安装有显存,可以不用占用系统内存,缺点是功耗大,发热量大

③核心显卡

(2) 打印机 按工作原理可分为击打式打印机(击打式针式打印机,又叫点阵打印机)和非击打式打印机(非击打式喷墨打印机和激光打印机)。

①针式打印机

②喷墨打印机

③激光打印机 特点是能输出分辨率很高且色彩很好的文字及图像,打字速度快、分辨率高、无噪音。

注意:衡量打印机质量的重要技术指标是分辨率,单位为每英寸打印点数(DPI)针式打印机在 360DPI 以上,喷墨打印机在 720DPI 以上,激光打印机可达 12000DPI 以上。

(3) 绘图仪

3.2.4 总线

1.总线的概念:是一组连接各个部件的公共通信线路,即两个或多个设备之间进行通信的路径。是计算机内部传输指令、数据和各种控制信息的高速通道,是计算机硬件的一个重要组成部分。

2.总线结构:

总线可同时传输的数据数就成为宽度,单位为比特。总线宽度越大,传输性能就越佳。总线的带宽=总线的工作频率 \times 总线的位宽/8。

3.总线的分类:

(1) 根据信号传送方式分类:

①并行总线:指的是数据的每位在各自的传输线上同时传送,互不干扰,一次传送整个信息。优点是传送速度快。但要求线数多,成本高,距离不远时可以采用

②串行总线:信息按顺序逐位传输,共享一条传输线。优点是只需要一条传输线,成本低,但是传送速度慢。

(2) 根据总线所在位置分类

①内部总线:是处理计算机内部各外围芯片与微处理器之间的总线,用于芯片一级的数据互通。

②系统总线:是计算机各扩展槽与主板之间的总线。

③外部总线:是主机与外部设备、计算机与计算机间连接的总线,用来实现和其他设备间的信息、数据交换。

(3) 根据传输的信息种类分类

① 地址总线：专门用来传送 CPU 发出的地址信息。由于地址只能从 CPU 传向外部存储器或 I/O 端口，即指明数据总线上数据的源地址或目的地址，地址总线总是单向三态的。其宽度决定了系统的最大存储能力，即 CPU 的寻址能力。

② 数据总线：在 CPU 与内存之间来回传送需要处理或需要储存的数据信息。是双向三态形式的总线。数据总线的位数是计算机的一个重要指标，通常与字长相一致。平常所说的计算机位数指的就是 CPU 数据总线的带宽。例如，如果数据总线带宽是 8 位，每条指令是 16 位，则处理器在一个指令周期内必须访问两次存储器模块。

③ 控制总线：用来控制数据总线和地址总线的访问和使用，即传送控制信号、命令信号和定时信号等。有的是 CPU 向内存或外部设备发出的信号；有的是内存或外部设备向 CPU 发出的信号。对每条控制线而言信号是单向传送，但作为整体是双向的（在各种结构框图中，凡涉及控制总线，均是以双向线表示）。
总线的标准：制定总线标准的目的是便于机器的扩充和新设备的接入。在计算机的发展中，CPU 处理能力的迅速提升促使了总线技术不断更新。

几种总线标准：

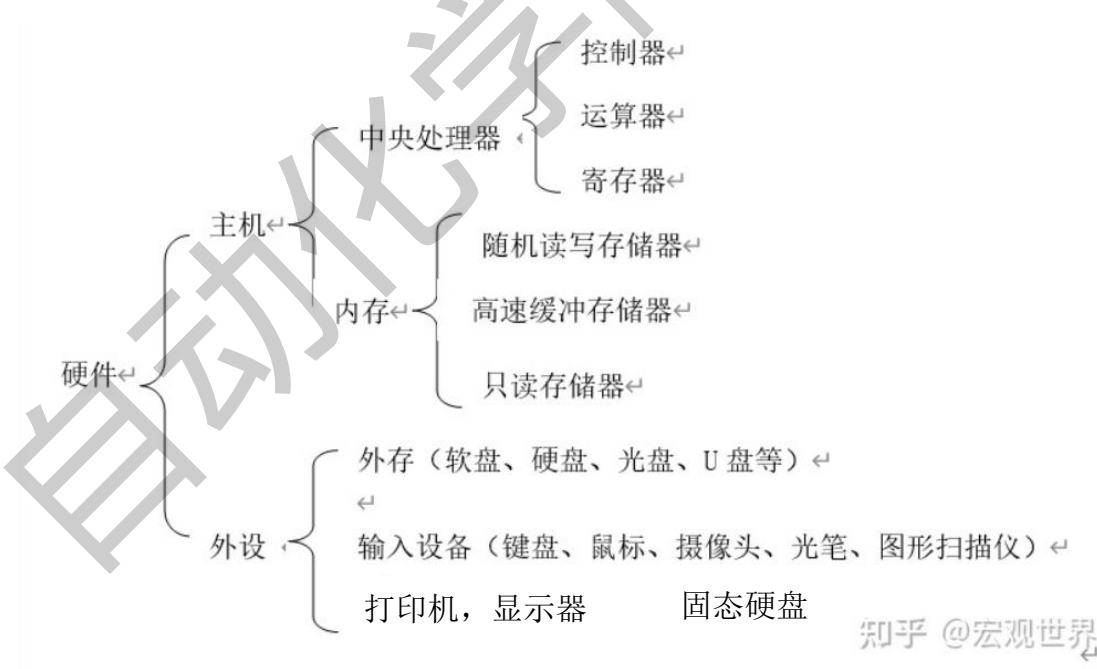
①ISA 总线：IBM 公司为了推行 PC/AT 机建立的系统总线标准也叫 AT 总线。

②EISA 总线：在 ISA 总线基础上扩充的开放总线标准。

③PCI 总线：目前广泛采用的一种高性能总线，1991 年由 Intel 公司提出。具有兼容性和可扩充性好、主板插槽体积小、支持即插即用等优点。

④USB 接口：是常用的总线标准，使用集线器经电缆分层形成树形结构，理论上最多可以连接 127 个外设。

⑤IEEE1394：是一种连接外部设备的机外总线，具有即插即用、支持带电热插拔、为外设提供电源等特性。从性能看最高传输率可达 480Mb/s



3.3 计算机软件

3.3.1 基本概念

1. 软件的定义：软件是指与计算机系统操作有关的计算机程序、规程、规则，以及相应可能有的文件、文档及数据。

完整的软件原则上由计算机程序、数据、文档和服务组成。

2.软件的特点:

- (1) 软件的抽象性
- (2) 软件的开发
- (3) 软件的复杂性
- (4) 软件的寿命: 不存在磨损和老化问题, 但存在缺陷维护和技术更新
- (5) 软件的设计: 大多是定制
- (6) 软件开发和运行的环境 必须依赖于特定的计算机系统环境
- (7) 软件研制成本

3.软件的分类: 详见书 P92

按照软件在计算机系统的功能划分, 可以将软件分为系统软件和应用软件两大类。

4.系统软件:

系统软件包括操作系统、高级语言处理系统、系统支撑软件、系统实用程序等。

- ①操作系统是系统软件的核心, 管理计算机硬件与软件资源的程序。
- ②高级语言处理程序包括汇编语言编译器, 高级语言编译、连接器等。
- ③系统支撑软件又称为软件开发环境, 主要包括环境数据库、各种接口软件和工具组。

5.应用软件: 在系统软件支持下开发提供的功能软件。如 WPS, EXCEL 等等

3.3.2 操作系统

操作系统的功能:

操作系统的功能包括作业管理、文件管理、处理机管理、存储管理和设备管理五个方面

操作系统的层次结构:

操作系统中定义了它的内核层和它与用户之间的接口。详见书 p95

3.3.3 从机器语言到高级语言

计算机语言可分为三类: 机器语言、汇编语言和高级语言

(1) 机器语言 (指令系统)

定义: 由计算机硬件系统可以识别的指令组成的语言。

构成: 指令 (操作码+操作数地址)

优点: 灵活、可直接执行、运行速度快。

缺点: 编写程序不方便, 容易出错, 调试查错非常困难。编写的程序只能在特定的机器上运行, 没有通用性。

指令系统: 指令集的体系, 计算机提供给用户的最原始的工具。

指令集称机器码, 也叫源代码。

(2) 汇编语言

定义: 机器指令映射成的一些可以被人读懂的助记符。

优点: 易于阅读理解且便于检查。

缺点: 要求用户熟悉计算机的指令系统, 可移植性差。

(3) 高级语言

定义: 以比较接近人类日常语言表述为基础的编程语言。

举例: FORTRAN、BASIC、COBOL、Pascal、C、C++、C#、Java、Python

将高级语言所写的程序翻译为机器语言的程序有如下两种:

① 编译程序

编译程序也称编译器,是采用生成性实现途径,把高级语言源程序(如 Fortran、Pascal、C 等)翻译成目标程序(机器语言程序)的一种程序,翻译的过程称为编译。

标准的编译系统,除基本功能外,还应具备语法检查、调试措施、修改手段、覆盖处理、目标程序优化、不同语言合用以及人机联系等重要功能。

②解释程序

解释程序也是一种对高级语言源程序进行翻译处理及的程序。但其处理方式是边读取、边翻译、边执行,解释过程不产生目标程序。解释程序将源程序一句一句读入,对每个语句进行分析和解释,有错误随时通知用户,无错误就按照解释结果执行所要求的操作。程序的每次运行都要求源程序与解释程序参加。

优点:简单易于实现,在解释执行过程中可灵活、方便地插入、修改调试措施,其方便性和交互性较好。

缺点:执行效率低。

3.4 软件工程

3.4.1 软件危机

软件产品通常具有复杂性、不可见性、易变性。

进行大型软件系统开发、运行和维护的过程中,遇到的规模越来越大,复杂程度越来越高,软件可靠性问题也越来越突出的情况称为“软件危机”。软件危机主要体现在以下几个方面:

- (1) 软件开发成本和开发进度难以预测
- (2) 用户对最终软件系统不满意的现象经常发生
- (3) 软件产品的质量往往不可靠
- (4) 软件产品难以维护
- (5) 软件缺少适当的文档资料

3.4.2 软件工程定义

软件工程是计算机领域的一个较大的研究方向,其内容十分丰富,包括理论、结构、方法、工具、环境、管理、经济、规范等,软件工程的最终目的是研究如何以较少的投入获得易维护、易理解、可靠性高的软件产品。

软件开发技术包括软件开发方法、软件工具和软件开发环境。良好的软件工具课促进方法的研制,而先进的软件开发方法能改进工具,软件工具集成软件开发环境。

软件工程管理可以保证软件开发质量。包括软件开发管理,软件心理学和软件经济管理。

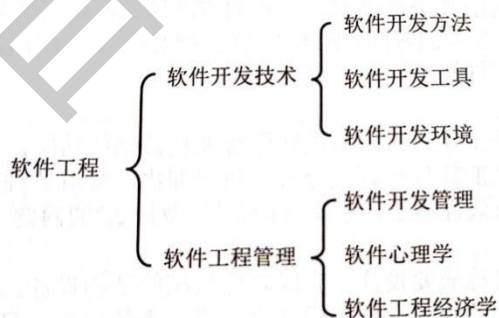


图 3-42 软件内容

3.4.3 软件生命周期

第4章 计算机方法学

4.1 问题求解

4.1.1 基本模式

1.发现问题

- ①依赖于个体对活动的态度。
- ②依赖于个体思维活动的积极性。
- ③依赖于个体的求知欲和兴趣爱好。
- ④取决于个体的知识经验。

2.明确问题

3.提出假设

4.检验假设

- ①实践检验
- ②间接验证方法

4.1.2 借助计算机的求解过程 详见书 p114

1.分析问题

2.设计算法

3.程序编码

4.1.3 两种问题求解过程的比较

人类求解问题的思路 and 过程与借助于计算机这一现代计算工具进行求解问题，还是存在着明显的差异。

(1) 在人类不一定需要建立对应问题的数学模型，主观性色彩较重。而借助于计算机技术进行求解问题时要先确定数学模型再进行计算处理，具有良好的合理性。

(2) 人类求解问题时，常表现出“心中有数”的状态。借助于计算机技术时则需要算法，及算法描述手段。

(3) 人类求解体现出针对问题的具体分析的能力。但大数据量的处理效率较低。相反计算机求解不再进行干预，体现出高效率，但在处理比人类“笨拙”、“刻板”。

(4) 在人类更擅长于对问题的形象思维（灵感），长时间重复容易疲劳和出错。而机器的特性使计算机擅长于确定的逻辑，明显的刻板 and 机械特性，但长时间重复不会出现疲劳 and 出错（除非硬件故障）。

4.2 问题的抽象

4.2.1 哥尼斯堡七桥问题

4.2.2 数学模型

数学模型建立的基本步骤如下：

对问题进行观察分析，明确建模目的；对问题进行必要合理的简化，搜集必需的各种信息，研究其运动变化情况；用精确的非形式语言进行问题描述，初步确定总的变量与相互关系；为了使处理方法简单，应尽量使问题线性化、均匀化。明确问题的系统种类、模型种类以及适当的描述数学工具。注意：工具越简单越有价值。

(1) 做出假设是建模至关重要的一步。

(2) 具体根据实际数据对数学模型进行测试、分析和统计，对模型解答进行数学上的分析。

(3) 检验修改模型是在问题的真实性与便于数学处理之间的折中过程。

汉诺塔问题

4.3 认识算法

4.3.1 什么是算法

1. 定义：算法就是为解决问题而采取的一组明确的、可以执行的方法或步骤的有序集合。解决问题的过程就是算法实现的过程。
2. 基本性质：确定性、可行性、输入、输出、有穷性。
3. 描述方法：自然语言、流程图、PAD 图、N-S 图、伪代码、计算机语言。
4. 评价：正确性、时间复杂度、空间复杂度、可读性、健壮性。

4.3.2 算法的描述

1. 自然语言描述

用人们日常使用的自然语言来描述或表示算法的方法，具体就是把算法的各个步骤，依次用人们所熟悉的自然语言表示出来。

使用自然语言描述算法的特点是通俗易懂。采用自然语言描述算法时，需要注意尽可能简洁、精确和详尽，避免产生歧义。

2. 图形化工具描述

(1) 流程图 详见书 p122。

(2) PAD 图

PAD 是用结构化程序设计思想表现程序逻辑结构的图形工具，它由程序流程图演化而来。也设置了 5 种基本控制结构的图式，并允许递归使用。

由于 PAD 的树形特点，使它比流程图更容易在计算机上处理。

详见页 p122。

(3) N-S 图

N-S 图是一种符合结构化程序设计原则的图形描述工具，又称盒图。

N-S 图的 5 种基本控制结构：详见书 p123。

N-S 图特点：

- ① 图中每个矩形框都是明确定义了的功能域，以图形表示，清晰可见。
- ② 它的控制转移不能任意规定，必须遵守结构化程序设计的要求。
- ③ 易确定局部数据和(或)全局数据的作用域。
- ④ 易表现嵌套关系和模块的层次结构。

3. 伪代码

伪代码是非正式的、半角式化、不标准的语言。采用伪代码描述算法结构清晰、代码简单、不拘于具体实现，可读性好。

伪代码不是用户和分析师的工具，而是设计师和程序员的工具。

4. 计算机语言

设计的算法也确实需要用“语言”恰当地表示出来。计算机无法识别和执行自然语言、流程图和伪代码形式描述的算法。这里涉及到大量的代码语言元素、语法规则和语言环境工具。采用计算机描述算法得到的结果既是算法也是程序，作为程序就可以直接上机运行算法。

4.3.3 算法的评价

1. 正确性

算法正确性是指算法设计应当满足具体问题的需求，它是评价一个算法优劣的最重要的标准。

2. 时间复杂度（运行时间）

算法的时间复杂度是通过算法编制的程序在计算机上运行时所消耗的时间来度量。算法的时间复杂度是一个问题规模 n 的函数 $f(n)$ ，它定量描述该算法的运行时间。

3.空间复杂度（占用空间）

算法的空间复杂度是指算法运行需要消耗的内存空间，主要包括包括存储算法本身所占用的存储空间，算法的输入输出数据所占用的存储空间和算法在运行过程中临时占用的存储空间这三个方面。

4.可读性

算法主要是为方便人们的阅读与交流，其次才是便于机器执行。算法的可读性好有利于人们对算法的理解；晦涩难懂的程序易隐藏较多错误，难以调试和修改。

5.健壮性

健壮性是指算法对不合理或非法数据输入的反应能力和处理能力，也称为容错性。它强调了即使输入了非法数据，算法应能够识别并做出正确处理，而不是产生莫名其妙的输出结果。

4.4 算法设计

4.4.1 穷举法

穷举法也叫枚举法，是通过把需要解决问题的所有可能情况逐一试验来找出符合条件的解的方法。

4.4.2 分治法

分治法采取“分而治之”的设计思想，通过降低问题规模达到快速解决问题的目的。问题特征有：

- （1）当问题的规模缩小到一定的程度就可以容易地解决；
- （2）问题可以分解为若干个规模较小的相同问题，即问题具有最优子结构性；
- （3）通过原问题分解出的子问题的解，可以合并获得该问题的解，这也是能否利用分治法的关键特征；
- （4）具体分解出的各个子问题需是相互独立的，即子问题之间不再包含公共的子问题，也是分治法效率的体现。

4.4.3 动态规划

动态规划算法先将问题拆分成多个简单的小问题，通过逐一解决这些小问题找到整个问题的答案。不同于分治法，动态规划拆分出的小问题之间相互关联。

动态规划采取的是分治法加消除冗余，是一种将问题实例分解为更小的、相似的子问题，并存储子问题的解而避免重复计算子问题，从而解决问题的算法。

4.4.4 递归法

递归算法是一种直接或者间接地调用自身算法的过程。

其具有以下特点：

- （1）递归就是在过程或函数里调用自身。
- （2）在使用递归策略时，必须有一个明确的递归结束条件，称为递归出口。
- （3）递归算法解题虽然很简洁，但解题的运行效率较低。所以一般不提倡用递归算法设计程序。
- （4）在递归调用的过程中，系统为每一层的返回点、局部量等开辟了栈来

存储。递归次数过多容易造成栈溢出，所以在实际编程中要注意栈溢出问题。

4.4.5 递推法

递推法是一种用若干步可重复的简单运算（规律）来描述复杂问题的方法。是序列计算中的一种常用算法。其思想是把一个复杂的庞大的计算过程转化为简单过程的多次重复。

递推法的解题步骤：

（1）按次序研究集合中最初最原始的若干问题。

（2）按次序寻求集合中问题间的转换规律即递推关系，使问题逐次转化成较低层级或简单的且能解决问题的或已解决的问题。

4.4.6 贪心法

贪心法是一种对某些求最优解问题的更简单、更迅速的设计技术，是一种最直接的方法。其特点是逐步进行，每一步骤中都采取在当前状态下最好或最优的选择，而不考虑各种可能的整体情况，从而希望导致结果是最好或最优的。

贪心法的设计策略：采用自顶向下，以迭代的方法做出相继的贪心选择，每做一次贪心选择就将所求问题简化为一个规模更小的子问题，通过每一步贪心选择，可得到问题的一个最优解。贪心法虽然每一步上都要保证能获得局部最优解，但由此产生的全局解有时不一定是最优的。

4.4.7 回溯法

回溯法是一种选优搜索法，按选优条件向前搜索，以达到目标。但当探索到某一步时，发现原先选择并不优或达不到目标，就退回一步重新选择，满足回溯条件的某个状态的点称为“回溯点”。

回溯法的基本行为是搜索，其求解问题的关键在于如何定义问题的解空间。回溯法的实现方法有两种：递归和递推。

4.5 程序设计基础

（1）基本概念

程序是为了实现特定目标或解决特定问题而采用计算机语言编写的命令序列的集合。Ps：程序与算法是有区别的，算法是程序的逻辑抽象，程序是精确的代码。

程序设计

是给出解决特定问题程序的过程。程序设计与程序编码不等同，程序编码一般在程序设计的工作完成后才开始。首先要先了解分析问题、设计确定解决问题的算法，再进行具体的程序编码。

程序设计方法 详见书 p136

（2）结构化程序设计

结构化程序设计的产生和发展形成了现代软件过程的基础。其主要特征：

①自顶向下：先考虑程序的整体结构，后考虑细节

②逐步求精：是指对复杂问题，设计一些子目标作为过渡，层层细化

③模块化：把一个大型的程序按照功能分解为若干相对独立的、较小的子程序(即模块)，并把这些模块按层次关系进行组织。限制使用 goto 语句 严格遵循每个程序段“只有一个入口和一个出口”的原则。

结构化程序设计三种基本结构：

①顺序结构②循环结构③条件选择结构

4.5.3 面向对象的程序设计

(1) 对象：对象不仅能表示具体的事物，还能表示抽象的规则、计划或事件。

对象实现了数据和操作的结合，使数据和操作封装于对象的统一体中。

(2) 属性：属性是用来描述对象的状态，对象的状态又称为对象的静态属性，包括对象内部所包含的信息，每个对象都具有自己专有的内部信息，这些信息说明了对象所处的状态。

(3) 类：类是对具有共同特征的对象们的进一步抽象，具有相同或相似性质的对象的抽象就是类。

类具有属性，属性是状态的抽象，类具有操作，它是对象行为的抽象。

(4) 消息：消息是向对象发出的服务请求，是对象和对象之间进行通信的手段。消息要素通常包括：发送对象、接受对象、操作和适当的参数。

面向对象程序设计具有以下的基本特性：

(1) 抽象性：抽象是人类认识问题的最基本手段之一。抽象是简化复杂的现实问题的途径。对一个问题的抽象包括两个方面：数据抽象和代码抽象。

(2) 封装性：封装目的在于将对象的使用者和对象的设计者分开。这种有效隐蔽和合理控制，就可以达到增强程序的安全性和简化程序编写工作的目的。

(3) 继承性：继承是表达类之间相似性的一种机制，即在已有的类的基础之上增加构造新的类，前者称为父类(或超类)，后者称为子类。

(4) 多态性：多态性是指在一般类中定义的属性或行为，被特殊类继承之后，可以具有不同的数据类型或表现出不同的行为。

多态性与继承性相结合使软件具有更广的重用性和可扩充性。

4.5.3 面向对象的程序设计

(1) 对象：对象不仅能表示具体的事物，还能表示抽象的规则、计划或事件。

对象实现了数据和操作的结合，使数据和操作封装于对象的统一体中。

(2) 属性：属性是用来描述对象的状态，对象的状态又称为对象的静态属性，包括对象内部所包含的信息，每个对象都具有自己专有的内部信息，这些信息说明了对象所处的状态。

(3) 类：类是对具有共同特征的对象们的进一步抽象，具有相同或相似性质的对象的抽象就是类。

类具有属性，属性是状态的抽象，类具有操作，它是对象行为的抽象。

(4) 消息：消息是向对象发出的服务请求，是对象和对象之间进行通信的手段。消息要素通常包括：发送对象、接受对象、操作和适当的参数。

5.2 数据库基础

5.2.1 基本概念

(1) 数据库

数据库(DB)是数据的集合，它具有统一的结构形式，存放于统一的存储介质内，并由统一机构管理。它由多种应用数据集成，并可被多个应用所共享，具有尽可能小的冗余度。

数据库存放数据，数据按所提供的数据模式存放，它能构造复杂的数据结构以建立数据间内在联系与复杂关系，从而构成数据的全局结构模式。

数据库中的数据具有“集成”、“共享”的特点，也即是数据库集中了各种应用

的数据，并对其进行统一的构造与存储，而数据可为不同应用服务与使用。数据库本身不是独立存在的，它是组成数据库系统的一部分。

(2) 数据库管理系统

数据库管理系统 (DBMS) 是统一管理数据库、使用户可以定义、创建和维护数据库以及提供对数据库有限访问的软件 (属系统软件)，DBMS 对数据库进行统一的管理和控制，以保证数据库的安全性和完整性。一般来说，DBMS 通过提供统一的数据语言具体完成：

①数据模式定义 ②数据操控 ③数据控制 ④数据交换 ⑤数据的扩展功能 ⑥数据服务 ⑦数据字典

(3) 数据库系统

数据库系统 (DBS) 是由数据库及其管理软件组成的系统。它是为适应数据处理的需要而发展起来的一种较为理想的数据处理系统，是一个为实际可运行的存储、维护和应用系统提供数据支撑的软件系统，是存储介质、处理对象和管理系统的集合体。

数据库系统由硬件系统、系统软件 (包括操作系统、数据库管理系统等)、数据库应用系统和各类人员等四部分组成。

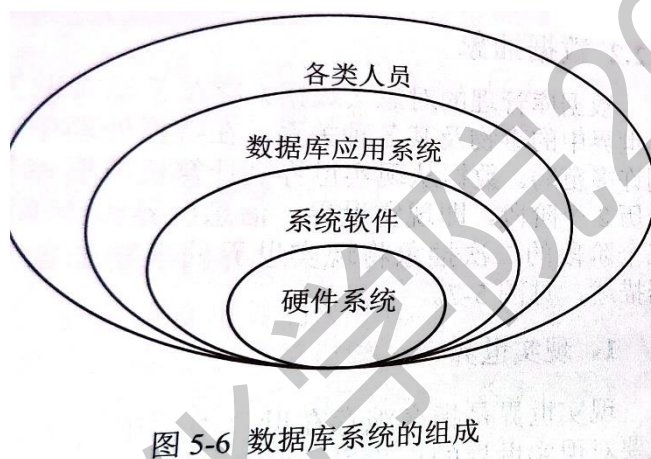


图 5-6 数据库系统的组成

5.2.2 数据抽象

数据库管理的对象 (数据) 存在于现实世界中，抽象于现实世界中的事物及其各种关系。在数据处理中，数据抽象涉及到许多范畴。数据从现实世界到计算机数据库里的具体表示要经历 3 个阶段，即现实世界、信息世界和计算机世界，并通过三个阶段的二次抽象将现实世界的事物抽象为计算机中的数据描述。

(1) 现实世界

现实世界是指客观存在世界中的事实及其联系。在这一阶段要对现实世界的事物进行收集、分类，并抽象成信息世界的描述形式。

(2) 信息世界

信息世界是现实世界在人们头脑中的反映。经过人脑的分析、归纳和抽象所形成的信息是对客观事物及其联系的一种抽象描述，把这些信息进行记录、整理、归类 and 格式化后，就构成了信息世界。在数据库设计中，这一阶段又称为概念设计阶段。

常用术语包括：

①实体。客观存在并可相互区别的事物。

②实体集。性质相同的同类实体的集合叫实体集。

③属性。实体的某一特征称为属性。

④元组。实体的每个属性都有一个确定值，称为属性的值。

⑤码。唯一标识实体的属性集称为码。

⑥联系。实体间的“联系”反映了现实世界中客观事物之间的关联。这种联系是复杂的、多种多样的，但归纳起来可分一对一、一对多和多对多三类。

(3) 计算机世界

这一阶段的数据处理是在信息世界对客观事物的描述基础上做进一步抽象，将其信息化，使得信息能够存储在计算机中，所以又叫数据世界，是对信息世界中信息的数据化。与信息世界常用概念对应，在计算机世界中涉及的术语：

①字段。对应与信息世界中的属性，用于标记实体属性的命名单位称为字段，或数据项。字段是数据库中可以命名的最小逻辑数据单位。

②记录。字段的有序集合称为记录。

③文件。同一类型记录的集合称为文件。

④关键字。能够唯一标识文件中每条记录的字段或字段集，称为关键字或主码。

从信息世界抽象到计算机世界是通过数据模型来表达的，在数据库系统中不仅要存储和管理数据本身，还要保存和处理数据之间的联系，也就是实体之间的联系，反映在数据上则是记录之间的联系，研究如何表示和处理这种联系是数据库系统的核心问题。

数据模型的选择，是设计数据库的首要任务，数据模型的好坏，直接影响数据库的性能。数据模型的设计方法决定着数据库的设计方法。常见的数据模型有层次模型（Hierarchical Model）、网状模型（Network Model）和关系模型（Relational Model）。

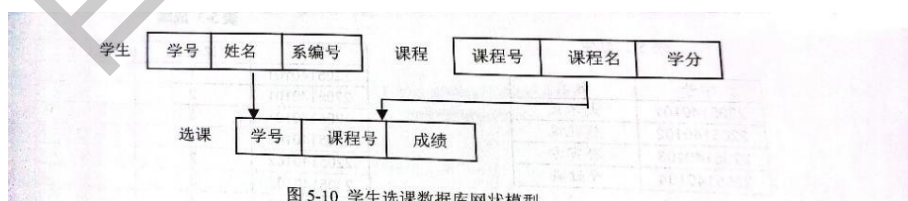
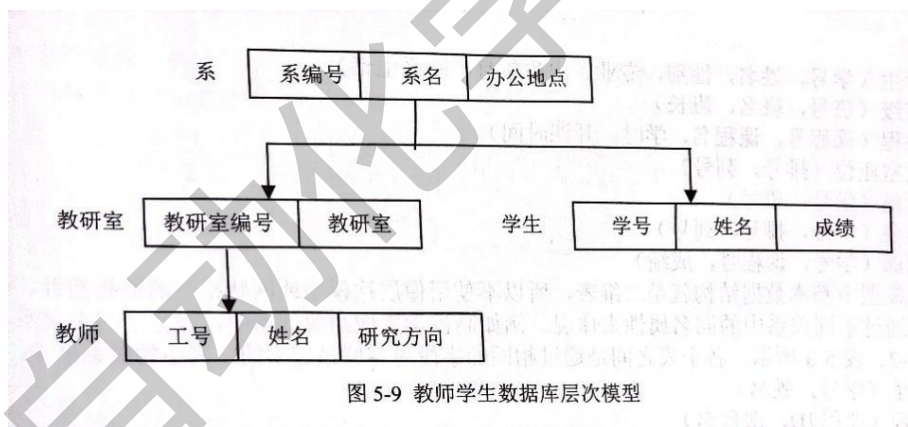


表 5-1 学生

学号	姓名
2205140101	赵治民
2205140102	钱礼悦
2205140103	孙寅方
2205140104	李虹威

表 5-2 课程

课程 ID	课程名称
1	高等数学
2	大学英语
3	军事理论
4	计算机基础

表 5-3 成绩

学号	课程 ID	成绩
2205140101	1	82
2205140101	2	78
2205140101	3	91
2205140102	1	72
2205140102	2	68
2205140102	3	69
2205140102	4	76
2205140103	1	75
2205140103	2	78
2205140103	3	81
2205140103	4	60
2205140104	1	89
2205140104	2	87
2205140104	3	75

现实世界、信息世界和计算机世界概念的对应关系如表 5-4 所示。

表 5-4 三个世界概念的对应关系表

现实世界	信息世界	计算机世界
事物个体	实体	记录
特征	属性	字段
事物总体	实体集	文件
唯一特征	实体标识符	关键字
事物间联系	E-R 模型	数据模型

1、三级模式结构：三级分别称为外模式、概念模式和内模式。

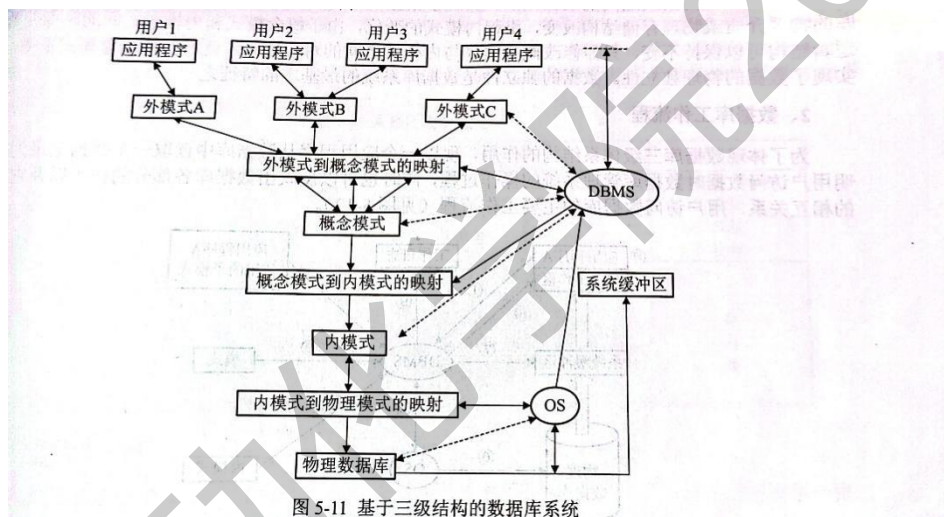
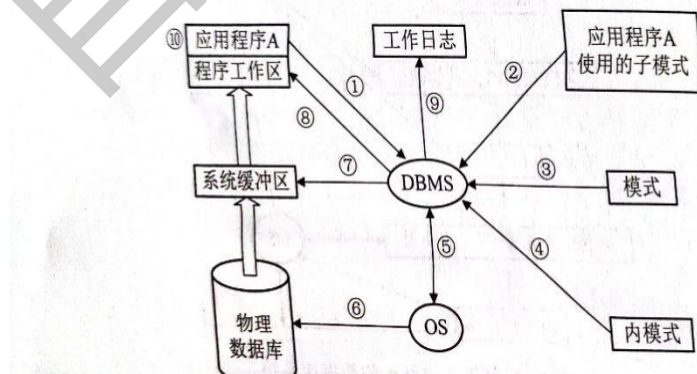


图 5-11 基于三级结构的数据库系统

2、数据库工作流程

DBMS 数据库工作流程：见书 P156



数据库及应用

常见的数据库:Access, XBase, SQL Server, Oracle 数据库, Mysql

新型数据库技术:①分布式数据库 ②并行数据库 ③多媒体数据库 ④数据仓库

数据库的典型应用:

- (1) 智能制造: 1.制造资源计划 2.企业资源计划 3.计算机制造系统
- (2) 空间数据库: 1.地理信息系统 GIS 2.数字城市 3.数字地球
- (3) 事务处理系统:
- (4) 处理作业方式分: 批处理系统和实时处理系统
- (5) 联机方式分: 联机集中式系统和联机分布式系统
- (6) 系统组织储存分: 使用文件或使用数据库的系统
- (7) 面向管理层次分: 高层, 中层, 操作层处理系统
- (8) 数据资源系统: 典型代表: 数字图书馆
- (9) 数据挖掘系统: <1>建立模型预测<2>分析关联<3>分类分析<4>聚类分析<5>序列分析<6>偏差检测<7>模式相似性挖掘<8>Web 数据挖掘

第 6 章 计算机网络

6.1 网络基础知识

6.1.1. 计算机网络的起源与发展

(1) 第一代计算机网络——面向终端的计算机网络阶段: 缺点: 主机负荷重, 通信线路利用率低, 网络结构集中控制方式, 可靠性低。

(2) 第二代计算机网络——计算机与计算机通信网阶段: 典型代表: ARPANET, 标志着计算机网络的发展进入了一个崭新时代。

(3) 第三代计算机网络——标准化网络阶段: ISO 委员会, 研究网络通信体系结构, 出现因特网。

(4) 第四代计算机网络——网络互联与高速网络阶段。

6.1.2. 计算机网络的定义与分类 a

定义: 计算机网络是利用通信设备和线路将分布在不同地理位置的、功能独立的多个计算机系统互连起来, 实现网络中资源共享和通信的系统。

1. 按照网络覆盖范围分类:

(1) 局域网: 在局部地区范围内将计算机、外设和通信设备连接在一起的网络。是最常见、应用最广的一种网络。特点: 连接范围窄、用户数少、配置简单、连接速率高。

(2) 城域网: 一般来说是在一个城市, 但不在同一地理小区范围内的计算机网络。特点: 连接距离远, 连接计算机数量更多, 但成本高。

(3) 广域网: 一般用来实现不同城市 LAN 或者 MAN 网络的互连, 多采用光纤线路。特点: 地理范围更广、总出口带宽有限。

局域网→城域网→广域网→互联网

2. 按照网络拓扑结构分类:

在计算机网络中, 人把计算机、服务器、交换机、路由器等网络设备抽象为“点”, 把网络中的电缆等通信介质抽象为“线”, 这样就可以将一个复杂的计算机网络系统抽象为由点和线组成的几何图形。人们称这种图形为网络的拓扑结构。网络拓扑结构的类型:

(1) 总线型网络

将所有的节点都连接到一条电缆上布线方式。设备：网卡、同轴电缆、接头 特点：结构简单、成本低、联网方便、易于扩充，但故障诊断较困难、实时性较差。说明：该拓扑结构基本被淘汰。

(2) 星型网络

由中央节点与各个计算机连接组成的网络。中心节点控制全网的通信。设备：网卡、双绞线、交换机。特点：结构简单、故障诊断容易、容易在网络中增加新的站点。缺点是中心节点负担较重。说明：该拓扑结构在局域网中应用最广。

(3) 环型网络

环型网络各个结点在网络中形成一个闭合的环，信息在环中作单向流动，可以实现任意两个结点之间的通信。特点：结构简单、路径选择方便，但可靠性差、网络管理复杂。说明：该拓扑结构基本被淘汰。

(4) 格型网络

格状网络结构中，结点之间的连接是任意的，没有规律，每个结点都至少要有两条线路和其他结点相连。其网络优点是可靠性高，即使一条线路出故障网络仍能正常工作，但网络控制和软件比较复杂，广域网基本上都是采用格状拓扑结构。其他分类：

科研网，教育网，校园网，企业网；专用网络和公用网络

3. 计算机网络性能指标

(1) 带宽：网络上能够同时传输信息的最大容量。

(2) 时延：指一个数据分组从网络的一端传输到目标端所需要的时间。

(3) 网络容量：指一个网络中所能容纳的最大的网络终端数目。

(4) 支持的协议与服务：网络所支持的协议越多，则能提供的服务也越多。

6.1.3 计算机网络功能：

数据传输，资源共享，分布式信息处理，提高可靠性

6.1.4 网络协议与体系结构

1. 网络协议三要素：语法，语义，同步规则。TCP/IP 是因特网标准协议

2. ISO OSI/RM 模型

OSI 是一个七层网络模型，包括物理层、数据链路层、网络层、传输层、会话层、表示层和应用层。

物理层：涉及网络接口和传输介质的机械、电气、功能和规程方面的特性。涉及的数据单位是二进制位（Bit）。

数据链路层：将不可靠的物理层转变成一个无差别的链路。分为介质访问控制和逻辑链路控制两个子层。涉及的数据单位是帧（Frame）。

网络层：负责将报文分组从源主机到目的主机的端到端传输过程。设计的数据单位是报文分组（Packet）。

传输层：负责整个报文从源到目的的传输，实现了高层与数据子网的隔离，关注的是报文的完整和有序问题。

会话层：负责网络会话的控制。比如会话的建立、维护和交互过程中的同步。

表示层：负责信息的表示和转换，比如发送加密，接收解密，中英文翻译等等。

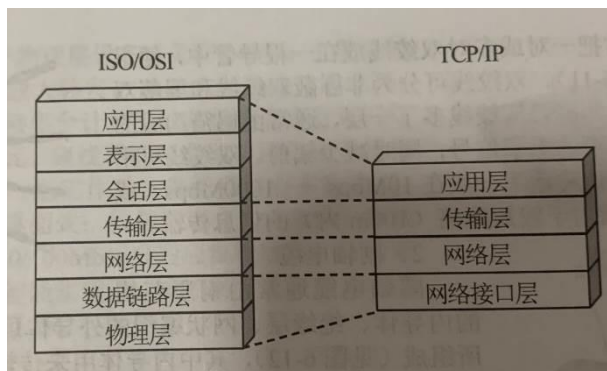
应用层：对应用程序的通信提供服务，负责向用户提供访问网络资源的界面。常用的电子邮件服务，文件传输服务等都是这一层提供的。

在 OSI 的 7 层模型中，网络层以下 3 层表现为通讯子网，之上的各层表现为资源子网。

注意：ISO OSI/RM 模型并不是一个具体的网络协议。

3.TCP/IP 体系结构

分为四层，网络接口层，网络层，传输层，应用层。



4.交换技术

电路交换：占用某一实际线路至数据交换结束。

报文交换：先发给中间设备，中间设备抽空发给目标。

分组交换：数据被划分为若干有严格上限的“分组”。由于能够保证同一用户不能长时间独占某一传输线路。适合交互式通信。

6.1.5 网络传输介质与关键设备

1.计算机网络传输介质

双绞线、同轴电缆、光导纤维、无线介质（微波、扩频无线电）

2.计算机网络关键设备

网络接口卡（网络适配器）：简称网卡，属于物理层设备；将结点与传输介质相连，完成数据接受和发送。

集线器：属于物理层设备；放大接受到的信号，扩大网络传输距离。

交换机：属于数据链路层设备；用于连接多个局域网

路由器：属于网络层设备，连接多个逻辑分开、使用不同协议和体系的网络或根据信道的情况自动选择和设定两个节点间的最近、最快的传输路径，并按先后顺序发送信号。

网关：连接不同软硬件的综合。

6.2 局域网

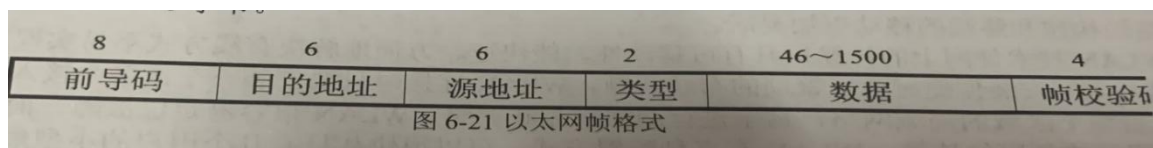
6.2.1 局域网标准：IEEE 802.X

6.2.2 以太网

（1）介质访问控制协议：总线型。用“发前先听，边发边听，冲突停止，延迟重发。

（2）以太网地址：MAC 地址，又称硬件地址，物理地址，第二层地址。厂商将 MAC 前 24 位买下作自己的标识符。后 24 位由厂商分配，生产时存入 ROM

（3）以太网帧格式：帧长固定，64 到 1518 字节。



(4) 典型以太网：10BaseT 以太网，交换式以太网，高速以太网。

6.2.3 无线局域网：

WLAN 灵活但会被障碍物阻碍电磁波传输且容易造成通信信息的泄露。目前最多使用第四第五代标准。

(1) wi-fi 详见书 P189

(2) 蓝牙 详见书 P190

6.3 因特网

6.3.1 6.3.2 详见书 P191-P196

6.3.3 IP 地址

因特网基本结构：分层互联结构

解决计算机间通信的问题的关键是通信协议，因特网采用的是 TCP/IP 协议。因特网是一个异构的计算机网络，凡是采用 TCP/IP 协议并且能够与因特网中的任何一台主机进行通信的计算机都可以看成是因特网的一部分。

TCP/IP 协议主要由传输控制协议（TCP）和因特网协议（IP）组合而成。IP 协议使计算机之间能够发送和接收分组，保证将数据从一个地址传送到另一个地址，但它不能解决传输中出现的问题。而 TCP 协议能解决传输中的问题，实现数据的可靠传输。

TCP/IP 是当前因特网协议簇的总称。该协议簇较为庞大，TCP 和 IP 是最重要的两个协议。

(1) 位于网络接口层的是各种物理网络的硬件设备驱动程序和介质访问控制协议，这些协议与物理网络相关，不在 TCP/IP 协议定义之列。

(2) 在网络层的底部是负责因特网地址（IP 地址）与底层物理网络地址之间进行转换的地址解析协议（ARP）和反向地址解析协议（RARP）。ARP 用于根据 IP 地址获取物理地址。RARP 用于根据主机的物理地址查找 IP 地址。IP 协议是网络层的核心协议，也是 TCP/IP 协议族中最为核心的协议，网络互连的基本功能主要由 IP 协议完成。

(3) 传输层只含 TCP 和传输层的用户数据报协议（UDP），两个协议提供进程间的通信，分别对应两类不同性质的服务。上层的应用进程可以根据可靠性要求或效率要求决定使用 TCP 或 UDP 提供的服务。

(4) 应用层的协议种类繁多，SMTP 支持电子邮件，HTTP 支持 WWW，FTP、

TFTP 和 NFS 支持文件的传输与访问，RIP 和 BGP 支持路由表的维护，TELNET 支持远程登录，BOOTP 和 DHCP 支持主机引导时自动获取信息，SNMP 支持网络管理等。

因特网是在网络级进行互联的，因此因特网在网络层（IP 层）完成地址统一的工作。IP 层用到的地址叫做因特网地址即 IP 地址。

因特网的 IP 协议提供了全因特网通用的地址格式（保证一致性），统一管理 IP 地址的分配（保证唯一性），以屏蔽物理网络地址的差异。

1. IP 地址结构

IP 地址由网络号和主机号两部分组成。网络号识别一个逻辑网络，主机号识别网络中的一台主机。相同的网络号属于一个逻辑网络。

网络标识的长度决定整个因特网能包含多少网络，主机标识的长度决定每个网络能容纳多少台主机。网络标识的长度不是固定的。

每台主机至少有一个 IP 地址，并且地址是全网唯一的。如果一台主机有多个 IP 地址。则该主机属于多个逻辑网络，称为多宿主主机，每个地址对应一个物理连接。由此，因特网地址本质是标识主机的网络连接。

目前广泛采用 IPv4，IP 地址由四个字节组成，地址空间为 2^{32} 。IP 地址采用十进制标记法，共 4 个数值，每个数值小于等于 255。中间用“.”隔开。

IP 协议定义了五类 IP 地址：A 类，B 类，C 类，D 类和 E 类。其中 A、B、C 是三个基本的类，分别代表不同规模的网络。

	第一字节	第二字节	第三字节	第四字节	第一字节取值	网络号	主机号
A 类	0				0~127	1 字节	3 字节
B 类	10				128~191	2 字节	2 字节
C 类	110				192~233	3 字节	1 字节
D 类	1110				224~239		
E 类	1111				240~255		

A 类地址：由 1 个字节的网络号和 3 个字节的主机号构成，用于少量的大型网络。0 一般不用，127 为回环地址，所以有 126 个可用网络。主机号全为 0 或 1 时不能用来表示主机，所以每个 A 类网络最多可容纳 $2^{24}-2$ 台主机，A 类地址第一个字节取值范围为 1~126。

B 类地址：由 2 个字节的网络号和 2 个字节的主机号构成，用于中等规模的网络。第一个字节最高 2 比特固定为 10，另外 14 个比特可变的网络号有 2^{14} 个。主机号全为 0 时表示网络地址，全为 1 时表示广播地址。这两个不能用来表示主机。所以每个 B 类网络最多容纳 $2^{16}-2$ （65534）台主机。第一个字节的取值范围为 128~191。

C 类地址：由 3 个字节的网络号和 1 个字节的主机号构成，用于小规模的网络。有 21 比特的可变网络号有 2^{21} 个，每个 C 类网络最多容纳 2^8-2 台主机。第一个字节的取值范围为 192~223。

D 类地址：用于组播。组播允许把消息传送给所有可能目的地中的一个经过选择的子集。D 类地址范围为 224.0.0.0~239.255.255.255，每个地址对应一个组，发往某一组地址的数据被该组所有成员接收。D 类地址不能分配给主机

E 类地址为保留地址，范围为 240.0.0.0~255.255.255.254

主机号部分全为 0 代表网络的 IP 地址。主机号部分全为 1 的表示某个网络上的所有主机（广播地址）。例如 192.168.10.0 是一个 C 类网络地址，192.168.10.255 表示该网络中的所有主机。

1. 子网及子网掩码

标准的 A 类、B 类、C 类网络可进一步划分为子网。原因有 5 点：

- (1) AB 类空间很大，难利用
- (2) 便于管理

- (3) 隔离广播通信, 减少堵塞
- (4) 出于安全考虑, 将管理与服务分开
- (5) 历史原因和应用需要, 便于互联一个单位的不同物理网络

划分方法: 将 IP 地址的主机号部分划分为两部分, 一部分表示子网, 一部分仍为主机号。

划分后 IP 地址可表示为 (网络号) (子网号) (主机号)

子网掩码指定了子网标识和主机号的分界点。子网掩码中对应于网络号和子网号的所有比特都被设为 1, 对应主机号设为 0。所以获取子网地址的方法是将子网掩码和 IP 地址进行按位“与”运算。

IP 协议:

IPv4: 地址资源有限, 枯竭且不平衡。

IPv6: 地址空间更大, 路由表更小, 增强组播支持, 增加自动配置支持, 实名化成为可能, 支持层次化网络结构, 具有更高安全性

6.3.4 IP 协议

IP 协议所处理的数据单元称为 IP 数据报。

IP 协议的一个核心任务是数据报的路由, 即决定发送数据报到目标机器的路径。路由选择围绕路由表进行, 进行路由选择的依据是网络的拓扑结构。每台主机和路由器中都有反应网络拓扑结构的路由表。

有关路由表操作的两部分: 一是根据路由表进行路由选择; 二是路由表的刷新及建立, 这项工作由路由守护程序完成, 它一般在系统引导时启动, 一直在后台运行。

数据分组在向信宿传递时分为直接传递和间接传递 (信源和信宿位于不同物理网络时) 两种方式

6.3.5 传输层协议: 详见书 P202

6.3.6 域名系统

1. 命名机制: 无层次命名机制和层次型命名机制

2. 因特网域名:

① 域名: 为方便记忆而为计算机进行命名。与 IP 地址作用同, 但便于记忆。(通常 95% 的个人用户入网后由 ISP 提供一个动态 IP 地址, 没有域名), 名字空间采用树形结构

② 域名系统 DNS: 完成域名向 IP 地址的转换。是由若干个域名服务器程序完成的。域名解析就是将域名翻译成 IP 地址的过程。DNS 将因特网划分为多个顶级域, 顶级域有组织模式和地理模式。顶级域下各管理组织划分二级域, 再将管理权授予下属管理机构继续划分。形成层次型域名结构。书写时, 顶级域放在最右面, 主机名放在最左面。

6.4 因特网服务及对人类的影响

6.4.1 因特网服务

1、电子邮件

- (1) 电子邮件 (E-mail), 又称电子信箱, 标志是 @。电子邮件可以包括超链接、HTML 格式文本、图像声音和视频等数据。
- (2) 电子邮件可以同时进行一对多的邮件传递, 具有快速传达, 不易丢失的特点。电子邮件不是一种“终端到终端”的服务, 而使用“存储转发式”的

通信服务方式,是一种异步通信。要收发电子邮件,必须有一个“邮箱”,邮箱用 E-mail 账号标识。E-mail 账号可向因特网服务提供商 ISP 申请,也可以通过某个网站申请。

- (3) 每个电子邮箱的 E-mail 地址都是唯一的,电子邮件地址的格式由三部分组成:
- (4) 邮箱名 @ 邮箱所在的邮件服务器的域名邮箱名是用户申请账号时指定的名字, @ 是分隔符, 域名一般就是用户注册的 ISP 的域名。
- (5) 电子邮件一般由邮件头和邮件体两部分组成。邮件头相当于邮件信封,包括多项信息,其中发件人的地址、发送的日期和时间等由系统自动生成,其他信息如收件人的地址等需要由发件人自行输入。
- (6) 邮件体是信件的具体内容,可以是文字信息,也可通过插入附件的形式传输图像、语音与视频等多种信息。

2、WWW 浏览

WWW(万维网),最初起源于欧洲粒子物理实验室,发明者为蒂姆·伯纳斯·李。

WWW 以超文本标记语言(HTML)和超文本传输协议(HTTP)为基础,采用客户机/服务器的工作模式,主要包括浏览器、Web 服务器和超文本传输协议三部分。

统一资源定位符 URL 也称 Web 地址,俗称“网址”,它规定了某一特定信息资源在 WWW 中存放地点的统一格式,即地址指针。

URL 的完整格式由以下基本部分组成:

协议+“:”+主机域名(IP 地址)+[:端口号]+目录路径+文件名

协议是指定与服务连接而使用的所有访问协议,协议类型也表示因特网资源类型。

主机域名指出 WWW 数据所在的服务器域名。

服务器提供的端口号表示客户可以访问服务器上不同资源类型。

目录路径指明服务器上存放被请求网页的路径。

文件名是客户访问页面的名称。

注意! WWW 上的服务器大多是区分大小写的,注意正确的 URL 大小写表达形式。

3、文件传输服务

文件传输服务由 TCP/IP 的文件传输协议(FTP)支持,通过网络将文件从一台计算机传送到另外一台计算机,是一种实时的联机服务,是因特网提供的基本服务之一。

基本功能:实现文件的上传和下载。另外,FTP 还提供对本地和远程系统的目录操作以及文件操作等功能。

FTP 以客户机/服务器方式工作。用户计算机运行 FTP 客户程序,称为 FTP 客户机,可申请 FTP 服务。远程运行 FTP 服务程序、并提供 FTP 服务的计算机称为 FTP 服务器,它通常是信息服务提供者的计算机。

因特网上的 FTP 服务器分为匿名 FTP 服务器和非匿名 FTP 服务器两类。

匿名 FTP 服务器是任何用户都可以自由访问的 FTP 服务器,用户使用“anonymous”作为用户名,用 E-mail 地址作为口令或输入任意的口令就可以登录。对于非匿名 FTP 服务器,用户必须首先获得该服务器系统管理员分配的用户名和口令,然后才能登录和访问,用户可以获得从匿名服务器无法得到的文件,还可上传文件到服务器。

访问 FTP 服务的方法有：浏览器，资源管理器，FTP 工具，FTP 搜索引擎。

4、P2P 技术

P2P 技术主要指由硬件形成网络连接后的信息控制技术，主要代表形式是在应用层上基于 P2P 网络协议的客户端协议的客户端软件。网络中所有的节点处于相同的地位，没有客户端和服务端之分，可以相互进行资源利用和数据共享。它一个重要的目标就是让所有的客户端都能提供资源。

P2P 技术有许多应用：

- (1) 对等计算。
- (2) 协同工作。
- (3) 基于 P2P 技术的搜索引擎，可以检索到网络上所有开放的信息资源
- (4) 用户可以利用基于 P2P 网络协议的客户端软件，直接从含有所需文件的结点机下载该文件。

5、Telnet

表示因特网远程登录服务的标准协议和主要方式。

主要用途表现在以下几个方面：

- (1) 远程登录缩短了空间距离。
- (2) 远程登录计算机具有广泛的兼容性。
- (3) 通过 Telnet 访问其它因特网服务。

注意：它将用户的所有内容都明文在互联网上传送，具有一定安全隐患，因此许多服务器都会选择禁用。

6、搜索引擎

指自动从因特网搜集信息，能够获得网站或网页的资料，经过一定整理以后，能够建立数据库且提供查询功能的系统。

全文搜索引擎是目前广泛应用的主流搜索引擎。（Google 和百度）

目录索引也称为分类检索，无需输入任何文字，只要根据主题分类目录层层点击进入。

6.4.2 网络空间安全

1、安全威胁

网络安全威胁是指某个人、物或事件对网络资源的机密性、完整性、可用性和非否认性所造成的危害。对网络的攻击可分为被动攻击和主动攻击。

被动攻击的目的是从传输中获得信息，其手段主要是对信息进行截获和分析，被动攻击分为析出消息内容和通信量分析。

主动攻击通常会比被动攻击造成的危害更严重，因为主动攻击通常要改动数据甚至控制信号，或者有意生成伪造数据。主动攻击可以分为篡改消息、伪装和拒绝服务。

2、网络安全技术

计算机网络安全服务中所采用的主要技术包括：密码技术、身份认证技术、数字签名技术、防火墙技术和防病毒技术等。

3.密码技术

密码系统通常可以完成信息的加密变换和解密变换。

通常按照在加密解密过程中使用的加密密钥和解密密钥是否相同，将密码体制分为对称密码体制和非对称密码体制。

常见对称密码算法有：DES、IDEA、RC5、AES 等。

常见非对称密码算法有：RSA、椭圆曲线密码算法、Diffie-Hellman 密钥交

换算法等。

4.身份认证技术

基本思想是通过对被认证对象的属性的验证来达到确认被认证对象是否真实有效的目的。

传统的认证技术主要采用基于口令的认证方法。

采用询问—响应方法。询问—握手鉴别协议采用的是询问—响应方法，它通过三次握手方式对被认证方的身份进行周期性的认证。用于远程拨号接入的点对点协议 PPP 给出了在点到点链路上传输多协议数据报的一种标准方法。

国际电信联盟的 X.509。

5.数字签名

数字签名是网络中进行安全交易的基础，目前正逐渐得到世界各国和地区在法律上的认可。我国的电子签名法已于 2005 年 4 月 1 日开始实施。

数字签名标准基于非对称密码体制，生成数字签名时使用私有密钥，验证签名时使用对应的公开密钥。只有私有密钥的所有者可以生成签名。

6.防火墙

防火墙可定义为：限制被保护网络与互联网之间，或其他网络之间信息访问的部件或部件集。防火墙的基本特性包括：

- (1) 内部网络和外部网络之间的所有网络数据流都必须经过防火墙。
- (2) 只有符合安全策略的数据流才能通过防火墙。
- (3) 防火墙自身应具有非常强的抗攻击免疫力。
- (4) 防火墙在互联网中是分离器、限制器、分析器。防火墙通常是一组硬件设备配有适当软件。
- (5) 防火墙在互联网中，对系统的安全起着极其重要的作用。但它还存在许多缺陷：
- (6) 防火墙为提高被保护网络的安全性，限制或关闭了很多有用但又存在安全缺陷的网络服务。
- (7) 防火墙无法防护内部网络用户的攻击。
- (8) 互联网防火墙无法防范通过防火墙以外的其他途径对系统的攻击。
- (9) 防火墙不能完全防止传送已感染病毒的软件或文件。所以在防病毒方面存在明显的缺陷。

7.入侵检测技术

入侵检测系统就是执行入侵检测任务的硬件或软件产品。它包括事件提取、入侵分析、入侵响应和远程管理四大部分，另外还可能结合安全知识库、数据存储等功能模块，提供更为完善的安全检测及数据分析功能。是一种用于检测任何损害或企图损害系统的保密性、完整性或可用性的网络安全技术。

入侵检测被认为是防火墙之后的第二道安全闸门，在不影响网络性能的情况下能对网络进行监测，从而对内部攻击，外部攻击和误操作的实时保护。