

# 汇编语言基本概念

## 计算机编程语言

**汇编语言：**助记符代替机器语言中的二进制操作码；地址符号或标号，代替指令或操作数的二进制地址。

汇编语言是不可移植的。汇编语言总是为特定系列的处理器设计的。

汇编语言主要用于优化程序中对执行速度要求苛刻的部分或实现对计算机硬件的访问。

**机器语言：**一种数字语言，只有计算机的处理器（CPU）才能理解。一条汇编语言指令对应一条机器语言指令。

**高级语言：**高级语言与汇编语言和机器语言是一对多的关系。

C++中也可以使用汇编，这就在使用高级结构和访问底层细节之间提供了一种折中的方案。用户可以用汇编语言直接访问硬件，但这将使得程序完全丧失可移植性。

表 1.1 汇编语言和高级语言的比较

应用程序类型	高级语言	汇编语言
用于单一平台的中到大型商业应用软件	正式的结构化支持使组织和维护大量代码很方便	最小的结构支持使程序员需要人工组织大量代码,使各种不同水平的程序员维护现存代码的难度极高
硬件设备驱动	语言本身未必提供直接访问硬件的能力,即使提供了也因为要经常使用大量的技巧而导致维护困难	硬件访问简单直接。当程序很短并且文档齐全时很容易维护
多种平台下(不同的操作系统)的商业应用软件 需要直接访问硬件的嵌入式系统和计算机游戏	通常可移植性很好,在不同平台可以重新编译,需要改动的源代码很少 由于生成的执行代码过大,执行效率低	必须为每种平台重新编写程序,通常要使用不同的汇编语法,难于维护 很理想,执行代码很小并且运行很快

## 虚拟机的概念

**虚拟机：**对于每个语言层次，将其想象成一台假想的计算机（虚拟机）。高层虚拟机的程序，通过解释或者翻译的方式，在底层虚拟机上执行。

## 虚拟机的层次

- 第5层：高级语言
- 第4层：汇编语言
- 第3层：操作系统
- 第2层：指令集体系结构
- 第1层：微结构
- 第0层：数字逻辑

## 执行方式

- 翻译方式：高层虚拟机的程序被**整体翻译**成底层虚拟机程序，然后在底层虚拟机上执行；
- 解释方式：低层虚拟机对高层虚拟机的程序，**逐条**指令进行**解码**并执行；

## 数据的表示方法

### 数据的存储单位

位（Bit）：表示一个二进制数码0或1，是计算机存储处理信息的最基本的单位。

字节（Byte）：一个字节由8个位组成。它表示作为一个完整处理单位的8个二进制数码。

字（Word）：16个位为一个字（即两个字节是一个字），它代表计算机处理指令或数据的二进制数位数，是计算机进行数据存储和数据处理的运算单位。通常称16位是一个字，32位是一个双字，64位是两个双字。

字长：字的位数叫做字长，不同档次的机器有不同的字长。例如一台8位机，它的1个字就等于1个字节，字长为8位。如果是一台16位机，它的1个字就由2个字节构成，字长为16位。

字节和字长的区别：由于常用的英文字符用8位二进制就可以表示，所以通常就将8位称为一个字节。字长的长度是不固定的，对于不同的CPU、字长的长度也不一样。8位的CPU一次只能处理一个字节，而32位的CPU一次就能处理4个字节，同理字长为64位的CPU一次可以处理8个字节。

双字（doubleword）：包含两个字

八字节（quadword）：包含两个双字

1 kB = 1024 byte

1 MB = 1024 kB

1 GB = 1024 MB

1 TB = 1024 GB

1 PB = 1024 TB

## 字符集

字符集是一个系统支持的所有抽象字符的集和。

### ASCII字符集

ASCII是一个7位的编码标准，编码的取值范围实际上是00h-7Fh。不同的计算机厂商对ASCII进行了扩充，增加了128个附加字符，它们的值在127以上的部分不是统一的。

### Unicode字符集

Unicode是ASCII字符编码的一个扩展，表示各种不同的国际语言。Unicode为每种语言中的每个字符设定了统一并且唯一的二进制编码，以满足跨语言、跨平台进行文本转换、处理的要求。

在Windows中用2字节对其进行编码，因此也被称为宽字符集(Widechars)。

Unicode是一种双字节编码机制的字符集，使用0~65535的双字节无符号整数对每个字符进行编码。

## 字节序

### 大端字节序

一个多字节组成的数据，最高位被存储在内存的低地址上。

127.0.0.1，对应的正整数16进制表示0x7F000001，大端字节序，表示为7F 00 00 01

网络中的通信数据使用大端字节序,也称为网络字节序。

### 小端字节序

一个多字节组成的数据，最低位被存储在内存的低地址上。

127.0.0.1，对应的正整数16进制表示0x7F000001，小端字节序，表示为01 00 00 7F

Windows程序中使用小端字节序。