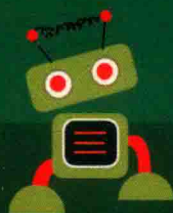
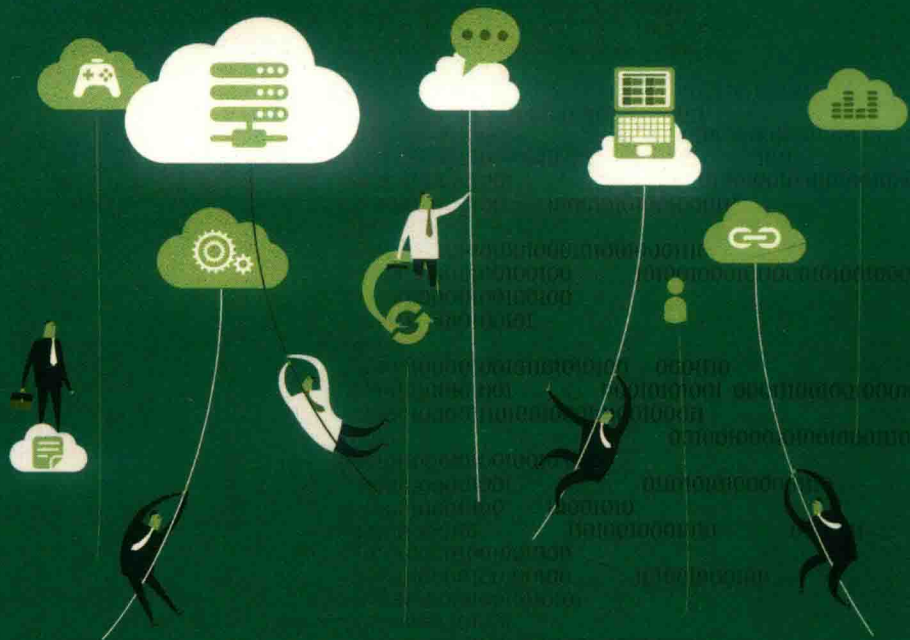


“新工科建设”教学成果

教育部大学计算机课程改革项目成果



大学计算机规划教材

# 数据与计算

(第4版)

◆ 陆汉权 编著

 中国工信出版集团




电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>

“新工科建设”教学成果  
教育部大学计算机课程改革项目成果  
大学计算机规划教材

# 数据与计算

(第4版)

陆汉权 编著



常州大学图书馆  
藏书章

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

本书以计算为主线，以数据及其表示为独特的视角，充分展示了计算机科学的全貌，包括其历史发展、技术特点、科学基础和未来前景。本书包括 9 章，主要内容为：计算系统，二进制和数字逻辑，数据表示，算法，编程语言和程序，数据库，网络与网络计算，大数据，先进计算。通俗的表达、生动的示例和大量的图表有助于读者对计算和数据相关抽象知识的理解。

本书可以作为高等院校大学计算机及相关课程的教材，也可以作为计算机和相关专业的计算机入门课程的教材。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有，侵权必究。

### 图书在版编目 (CIP) 数据

数据与计算/陆汉权编著. —4 版. —北京: 电子工业出版社, 2019.9  
ISBN 978-7-121-36701-4

I. ① 数… II. ① 陆… III. ① 计算机科学—高等学校—教材 IV. ① TP3

中国版本图书馆 CIP 数据核字 (2019) 第 106666 号

责任编辑: 章海涛

印 刷: 北京虎彩文化传播有限公司

装 订: 北京虎彩文化传播有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1 092 1/16 印张: 14.5 字数: 371 千字

版 次: 2011 年 8 月第 1 版

2019 年 9 月第 4 版

印 次: 2019 年 10 月第 4 次印刷

定 价: 45.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 [zltz@phei.com.cn](mailto:zltz@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式: 192910558 (QQ 群)。

# 前 言

超级计算机在新闻报道中常常被简称为“超算”。根据 TOP500 榜单定期发布的世界最快的超级计算机排行榜，我国的天河系列（2009 年）和神威（Sunway，2016 年）超级计算机曾经先后占据世界第一的位置累计 5 年之久，而目前最新的排行榜中位列第一的是美国 IBM 的 Summit。神威是世界上第一台运算速度超过 10 亿亿次的超级计算机，Summit 则是第一台运算速度超过 20 亿亿次的超算机器。中美两国都在研发 100 亿亿次的超算机器。神威和 Summit 这类运算速度极快的机器，能够解决过去被认为不可能计算的复杂任务，也能够模拟自然环境、探知未来世界和宇宙空间等。中美在网络技术发展方面也在逐力，华为公司被美国列入“实体名单”，很大原因在于华为是第 5 代移动数据网络技术的领先者。不可否认，数据网络原创地是美国，但中国作为后来者已经在应用方面成了事实上的领先者。不仅如此，人工智能将是接下来角逐的领域之一。

上述现实也充分说明了人类社会发展到数字时代，国家之间的竞争已经进入了非传统领域。无论主观意愿如何，我们都将融入数字世界。显而易见，仅仅会使用计算机和手机是远远不够的。要能够从容应对这个被计算机主导的技术化社会，就需要有相关的、宽泛的知识背景，而不是简单地会上网和使用流行的应用软件。

本书前后有多个版本，书名从最初的《计算机导论》到今天的《数据与计算》，也蕴含着计算技术的发展从计算机到深层次的应用，如大数据、智能计算等。另一方面，本书的主要读者是大学新生，他们都是计算机和互联网的“原住民”。他们诞生之时，互联网已经开始快速发展，而计算机特别是个人计算机早就是最流行的消费类电子产品了。从这个意义上说，他们更应该去理解信息社会特征，掌握计算机科学的基础知识。即使单纯将计算机视为工具的读者，也肯定明白这个道理：用好工具的前提是了解工具的原理。

本书力求具有较好的可读性，尤其考虑到很多读者是非计算机专业的学生，他们今后的专业学习也将与计算机密切相关。毫不夸张地说，无论哪个专业，计算机都是不可或缺的学习、研究的重要工具，也是工作和生活的重要帮手。

本书从 2015 年开始以“数据与计算”为名，就是突出其数据组织和表达的抽象特点，介绍建立在数据之上的各种计算过程、方法，以帮助“原住民”们更好地理解计算机这个机器的计算能力。

本书安排的主题是从计算机开始。其实这并不是新概念，恰恰是“旧观念”：计算机的重要贡献者冯·诺依曼就是以“数字计算机”描述计算机的。第 1 章介绍的是具体的计算机。在计算机科学中，计算机这个硬件装置可能是唯一能够被触摸的实体。从实体机器到抽象的数据概念，这样有利于理解计算，有利于理解数据的抽象表达。因此第 2 章就自然过渡到计算需要的数，引出了二进制和相关的二值逻辑表达，介绍逻辑关系实现了算术计算。二进制和数字逻辑是计算机科学的最重要的基础内容。计算机中的所有数据都是二进制的，如何使用二进制数表示各种不同的对象，如文本、文档和多媒体数据，即为第 3 章所介绍的内容。

任何一种数据表示都需要有相关的算法，第 4 章会介绍算法的基本知识。算法的概念很重要，社会学家把人分为“数据提供者”和“算法设计者”，其中的内涵不言而喻。算法有趣、

有用，也深刻地揭示了计算机的本质就是按照人类预先设定好的步骤去完成任务。算法源于数学，已经成为计算机科学的基石。如果把编程语言和算法集合起来，就是今天成千上万的应用程序。第5章介绍了编程语言和程序，这里充分展示出计算抽象的另一面：形式化表达。尽管算法和程序的逻辑关系很简单，但由简单到复杂的算法和程序的设计过程，也充分展示出人类聪明才智的巅峰：大型程序的复杂性堪比人类已知的复杂工程。

本书后续几章都是围绕应用展开的：第6章的数据库、第7章的网络与网络计算、第8章的大数据、第9章的先进计算。超算、人工智能和计算理论、计算复杂性等作为本书的内容，也留下了很多待解的疑问。未来有很多是不确定的，在计算领域也是如此。

相关计算机安全、虚拟化世界等问题，本书在第7章中进行了介绍。尤其是计算机与网络安全、隐私、数据保护等相关技术和社会问题，这些都是“原住民”们应该知道其技术背景和进一步思考其深远影响的大问题。

最后是写给本书的主要读者——大学新生的。

首先，你们需要学习的东西远远超过书本上提供的。你们已经熟悉如何从网络中获取知识，但需要了解这些“知识”的来源及其正确性、真伪性。

其次，理解知识比了解知识更重要，理解计算机、理解信息、理解数据，不仅仅是概念，重要的是它们对你的影响。设想，你是否可以一整天不使用手机、不使用计算机？

再次，学习很重要，你们需要学会学习。书中的很多名词、术语真的难记，不要紧，理解即可。知识需要积累，只有积累了足够的知识，才能重构你的思维，才能从已有的知识中发现解决问题的新方法。本书提出了很多问题，有些有意没给出解答，有的解答也可能是含糊的，也许这些问题是没有答案的。希望你们能够认真地低头看书，然后抬头思考。有很多问题值得追问“为什么”。

感谢您阅读或使用本书，希望本书没有让您失望。也希望读者能够发现和纠正书中的各种错误，无论是观点错误，还是概念错误，包括错别字，作者都真诚地欢迎并接受，并将在今后的重印或修订中予以纠正。

感谢电子工业出版社长期以来对本书的鼎力相助。

感谢作者所任教的浙江大学计算机学院的很多老师。有他们对本书的指导和鼓励，作者才有写作本书的动力。

陆汉权

2019年6月于浙江大学

# 目 录

第1章 计算系统	1
1.1 计算机系统	1
1.1.1 硬件	1
1.1.2 软件	3
1.2 计算机简史	3
1.2.1 硬件史	3
1.2.2 软件史	4
1.3 计算机模型	5
1.3.1 数据处理机模型	5
1.3.2 现代计算机模型	6
1.4 计算机组成	8
1.4.1 CPU	8
1.4.2 存储器	9
1.4.3 输入和输出	12
1.5 操作系统	13
1.5.1 计算机的核心	13
1.5.2 资源调度	14
1.5.3 文件系统	17
1.6 计算机网络	18
1.7 数据和信息	19
本章小结	20
第2章 二进制和数字逻辑	26
2.1 数据的表示	26
2.2 数制	27
2.3 二进制数	29
2.4 二进制运算	30
2.5 数字逻辑	33
2.5.1 基本逻辑关系	33
2.5.2 逻辑代数	34
2.6 逻辑电路	36
2.6.1 门电路	36

2.6.2	加法器 .....	37
2.6.3	存储单元电路 .....	38
2.6.4	集成电路 .....	39
本章小结	.....	40
习题2	.....	41
<b>第3章</b>	<b>数据表示</b> .....	<b>46</b>
3.1	数据概述 .....	46
3.2	文本和文档 .....	47
3.2.1	ASCII .....	48
3.2.2	Unicode编码 .....	48
3.2.3	汉字编码 .....	49
3.2.4	文档 .....	50
3.3	数据压缩 .....	50
3.3.1	霍夫曼编码 .....	51
3.3.2	行程长度编码 .....	52
3.3.3	有损压缩 .....	52
3.4	音频数据 .....	53
3.4.1	音频数据采集 .....	53
3.4.2	音频数据格式 .....	55
3.4.3	计算机语音 .....	56
3.5	图形和图像 .....	57
3.5.1	图像的表达 .....	57
3.5.2	3D技术 .....	60
3.6	视频和动画 .....	61
3.6.1	视频数据 .....	61
3.6.2	动画 .....	62
本章小结	.....	63
习题3	.....	64
<b>第4章</b>	<b>算法</b> .....	<b>68</b>
4.1	算法概述 .....	68
4.2	算法的三种结构 .....	69
4.3	算法的表示和发现 .....	71
4.4	算法举例 .....	73
4.4.1	基本算法 .....	73

4.4.2	迭代 .....	75
4.4.3	递归 .....	75
4.4.4	排序 .....	77
4.4.5	查找 .....	78
4.4.6	搜索图 .....	79
4.5	算法的方法学 .....	80
4.5.1	贪心法 .....	80
4.5.2	分治法 .....	82
4.5.3	动态规划 .....	83
4.5.4	回溯法 .....	85
4.6	抽象数据表达 .....	87
	本章小结 .....	88
	习题4 .....	89
<b>第5章</b>	<b>编程语言和程序 .....</b>	<b>93</b>
5.1	程序概述 .....	93
5.2	编程语言 .....	94
5.2.1	机器语言和汇编语言 .....	94
5.2.2	面向过程的高级语言 .....	95
5.2.3	面向对象的程序设计语言 .....	97
5.2.4	其他语言 .....	99
5.3	程序的程序：翻译系统 .....	100
5.4	高级编程语言 .....	101
5.4.1	数据类型 .....	101
5.4.2	运算操作 .....	103
5.4.3	基本语句 .....	105
5.4.4	控制语句 .....	106
5.4.5	函数和方法 .....	107
5.5	程序编写 .....	108
	本章小结 .....	110
	习题5 .....	111
<b>第6章</b>	<b>数据库 .....</b>	<b>114</b>
6.1	数据库概述 .....	114
6.1.1	文件管理 .....	114
6.1.2	数据库方法 .....	115

6.1.3	数据库管理系统 .....	116
6.2	关系数据库 .....	117
6.2.1	关系 .....	117
6.2.2	关系运算 .....	119
6.2.3	SQL .....	120
6.2.4	事务 .....	122
6.3	非关系数据库 .....	122
6.4	其他数据库技术 .....	124
6.5	构建信息系统 .....	125
6.5.1	数据库设计 .....	125
6.5.2	访问接口和数据转换 .....	126
6.5.3	访问结构 .....	127
	本章小结 .....	128
	习题6 .....	129
<b>第7章</b>	<b>网络与网络计算 .....</b>	<b>132</b>
7.1	通信基础 .....	132
7.1.1	调制解调 .....	132
7.1.2	网络介质 .....	133
7.1.3	无线通信 .....	134
7.1.4	带宽 .....	136
7.1.5	压缩和校验 .....	137
7.2	网络技术 .....	137
7.2.1	网络访问 .....	138
7.2.2	网络类型 .....	139
7.2.3	网络硬件 .....	141
7.2.4	网络协议 .....	142
7.3	互联网 .....	143
7.3.1	TCP/IP .....	144
7.3.2	互联网基本概念 .....	146
7.3.3	万维网Web .....	149
7.3.4	互联网服务 .....	152
7.3.5	5G和物联网 .....	154
7.4	网络数据 .....	155
7.5	云计算 .....	158
7.6	网络安全 .....	160

7.6.1	病毒和黑客 .....	160
7.6.2	隐私保护 .....	162
7.6.3	密码保护 .....	163
7.6.4	电子签名 .....	165
本章小结	.....	166
习题7	.....	168
<b>第8章</b>	<b>大数据</b> .....	<b>173</b>
8.1	大数据概述 .....	173
8.2	R简介 .....	174
8.3	大数据预处理 .....	176
8.4	数据分析方法 .....	177
8.4.1	聚类分析 .....	177
8.4.2	关联分析 .....	179
8.4.3	回归分析 .....	182
8.5	数据挖掘 .....	185
8.5.1	分类 .....	186
8.5.2	时间序列分析 .....	188
8.5.3	文本分析 .....	190
8.6	大数据处理工具 .....	192
8.6.1	Hadoop .....	192
8.6.2	MapReduce .....	193
本章小结	.....	194
习题8	.....	195
<b>第9章</b>	<b>先进计算</b> .....	<b>197</b>
9.1	高性能计算 .....	197
9.2	人工智能 .....	198
9.2.1	图灵测试 .....	198
9.2.2	强人工智能和弱人工智能 .....	199
9.2.3	哪些问题需要智能计算 .....	200
9.2.4	启发法 .....	201
9.2.5	知识表达 .....	202
9.2.6	神经网络 .....	203
9.2.7	机器人 .....	204
9.2.8	自然语言处理 .....	205

9.2.9	人工智能算法简介 .....	206
9.3	机器学习和深度学习 .....	207
9.3.1	机器学习 .....	208
9.3.2	深度学习 .....	208
9.4	虚拟现实 .....	209
9.5	计算理论 .....	210
9.5.1	图灵机 .....	210
9.5.2	停机问题 .....	212
9.5.3	P和NP问题 .....	213
	本章小结 .....	215
	习题9 .....	215
附录 A	ASCII 表 .....	218

# 第 1 章 计算系统

计算系统包括两部分：计算机和数据。这个概念是由冯·诺依曼（John von Neumann, 1903—1957）于 1945 年提出的。冯·诺依曼在计算机和博弈论方面都有杰出的贡献。计算系统是一个极其复杂的系统，其复杂性被飞速发展的技术所隐藏，使得计算机得以普遍应用。本书的目的是解释这个复杂系统。本章从介绍计算机开始，通过计算机模型、计算机硬件、操作系统、数据与信息、网络等基本概念和重要术语的介绍，以期读者对计算系统有一个初步的、整体的认识。

## 1.1 计算机系统

如果问“计算机是什么”，可能每个人都有自己的答案。如果问“计算机能做什么”，可能的回答是无所不能。当然，真实的情况并非如此。还有一个问题，“计算机里有什么”。回答这个问题就有点难度了。只要不是计算机专业人员，哪怕是对计算机（智能手机）熟悉的很多人，他们频繁使用很多应用，如网络购物、外卖、打车、订票、发微信、地图定位、视频通话等，但不一定了解计算机究竟是如何运作，不一定了解计算机的基本知识。

不知道计算机原理，并不影响人们使用它。但接受良好的教育包括学习计算机的基本知识，了解计算机技术如何影响着社会，如何影响着人类的生活，网络如何改变着世界。同时，理解计算机处理问题的方法和过程也是一种科学素质，有助于个人能力的提升。

计算机的核心技术领域有硬件、软件和网络通信。就计算机这个机器而言，它是一个系统，由硬件和软件组成，如表 1-1 所示。

本节介绍计算机系统的基本知识。有关计算机的组成部件在 1.4 节中进一步介绍。

许多文献中使用“计算机”作为其正式名称，但广为人知的名称是“电脑”。本书对这两个名词不加区别地使用。

表 1-1 计算机系统的组成

计算机	硬件	处理器（主机）
		存储器
		输入/输出设备
	软件	系统软件
		应用软件（App）

### 1.1.1 硬件

实际上，计算机（computer）的原意是“从事计算的人”。计算机的重要贡献者冯·诺依曼最初将计算机称为“自动计算系统”，后来以“电子通用数字计算机”命名。在本书中，“计算机”或“计算机系统”多与机器相关，而计算系统指计算机和数据。

计算机硬件 (hardware) 是指物理装置的机器。图 1-1 所示的是不同种类的计算机, 从左到右分别是服务器 (server)、台式一体机 (desktop, 主机和显示器被制造在一起)、笔记本电脑 (notebook/laptop)、平板电脑 (pad) 和智能手机 (smartphone)。服务器多为企业或机构使用, 其余是个人计算机。还有更大体积的计算机——超级计算机, 图 1-2 是无锡国家超级计算机中心的神威太湖之光 (Sunway Taihulight), 由 40 个机柜 (体积如同较大的双门冰箱) 和 8 个网络机柜组成。神威太湖之光从 2016 年 6 月开始占据世界最快计算机的位置达两年多 (最快计算机排行榜: [www.top500.org](http://www.top500.org))。2018 年 11 月, 排名首位的是美国 IBM 公司的机器。无论大小, 这些计算机都被称为计算机系统 (computer system)。

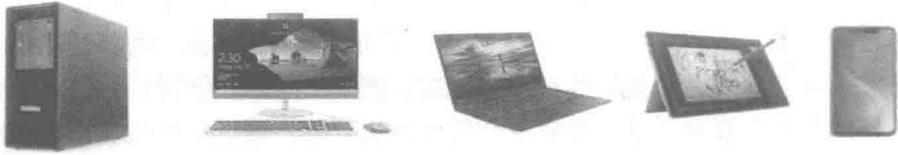


图 1-1 不同种类的计算机

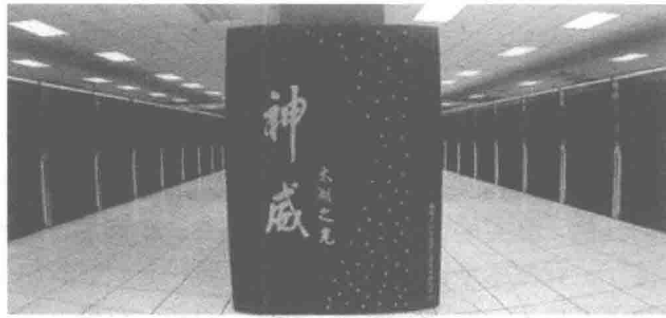


图 1-2 神威太湖之光超级计算机

计算机分类没有标准, 最简单的方法是按计算机的规模及售价划分: ① 数千万美元的超级计算机 (supercomputer), 如神威太湖之光。这类机器体积庞大, 价格昂贵, 通常用于极为复杂的、海量的数据处理领域, 如地震模拟、核武器试验、气象分析、生物信息处理等。② 数百万美元的大型计算机 (mainframe computer), 多用于跨国企业的信息系统, 如银行、航空公司等。③ 数万至数十万美元的小型计算机 (small computer), 一般用作企业、政府、学校等机构的网络服务器, 或用于研究机构、较大型的工程设计领域。④ 价格便宜的微机 (microcomputer), 用户最多。更为人熟知的微机的另一个名字是 PC (Personal Computer, 个人计算机), 是美国 IBM (International Business Machines) 公司 1981 年生产的微机的商标。此后, PC 就成为了微机的代名词。笔记本电脑 (notebook computer) 早期的名字是膝上机 (laptop), 是一种便携式的微机。

比微机更小的是嵌入式计算机。今天的大多数实验室仪器、医疗设备、工业设备都带有数据处理能力, 这种处理能力依赖于嵌入其中的计算机芯片。家用电器, 如电视机、电冰箱等, 也内嵌计算机芯片, 以实现更高级的功能。交通工具, 如飞机、轮船和汽车, 不但使用计算机进行设计、制作, 而且其内部嵌入了大量的计算机芯片实现自动检测、控制和导航等功能。

另一类是移动设备, 如平板电脑、智能手机等, 有时也将笔记本电脑归类为移动设备。智能手机 (smartphone) 就是附加了电话功能的掌上机, 具有个人计算机的大部分功能。当然, 这里的智能只是指“计算机功能”, 还没有真正意义上的智能。一直在研究之中的可穿戴式计

算机，虽然没有形成真正意义上的产品，但某些带有处理、通信的类似玩具的智能手环、智能眼镜等也开始问世，并被热捧。

## 1.1.2 软件

计算机的硬件是看得见的物理部分，而计算机的软件（software）是看不见的。软件是控制计算机硬件运行的程序。硬件与软件的关系有一个很好的比喻：如果计算机硬件是躯体，那么软件就是灵魂。实际上，计算机硬件之外的所有东西，包括文档、程序、语言等，都被归类为软件。因此，软件也是一个极其复杂的系统。

软件分为系统软件和应用软件两类。服务于计算机本身的那些软件称为“系统软件”（system software），其主要任务是确保计算机能够有效地完成用户的任务。系统软件包括操作系统（operating system），如 Windows、MacOS/iOS、Android（安卓）等，也包括计算机编程语言，以及硬件检测和管理的一些工具软件。

软件的另一类是应用软件（application software），或者称为应用程序，现在常以“App”（读[æp]）代之。尽管从专业术语上对程序（program）和软件的定义是不同的，但很多人将 App 视为程序、软件的代名词。

应用软件是解决特定应用问题的软件。例如，字处理软件、电子表格、演示软件、电子邮件等是通用型的应用软件。很多计算机应用都需要专门的软件，如学校的教务系统就是为了进行学生选课、课程安排、成绩记载、学籍管理等工作的应用软件。

从数据的角度看，程序对数据进行处理是完成计算功能。这里的计算泛指处理过程，如文字处理、事务处理等。被冠以计算的很多术语，如智能计算、云计算、边缘计算，就是因为它们是基于计算机的，当然需要相应的程序。从计算系统的角度看，程序本身也是数据。

## 1.2 计算机简史

在 20 世纪 40 年代的第二次世界大战期间，为破译通信密码和解决新型火炮弹道的复杂计算，美国开始研制计算机。相比于数学、物理、化学等学科，计算机学科的历史很短，但对人类社会的影响确实无与伦比：计算机颠覆性地改变了人类的生活、工作、学习方式，改变了商业模式和生产制造过程。

人类使用计算工具的历史可以追溯到数百年前，如我国的算盘就是一个古老的计算工具。一般说来，20 世纪 40 年代前的各种计算工具，如算盘、手摇计算机、机械式计算机等，都不被认为是计算机，因为这些装置不能实现“自动计算”。人们把 1946 年的 ENIAC（Electronic Numerical Integrator And Computer，电子数字积分计算机）称为第一台（自动）计算机。下面从硬件和软件两方面简单介绍计算机的历史。有兴趣的读者可访问计算机历史博物馆网站：<https://www.computerhistory.org/>，其中介绍得很详细。

### 1.2.1 硬件史

计算机问世的初期，计算机就是指计算机的硬件。由于硬件是计算机的物理体现，因此在计算机发展史中，主要依据其硬件的技术特征作为标志，并没有严格的时间划定。

ENIAC 是第一代计算机（1945—1954）的典型代表，采用电子管作为主要器件。这个阶段后期，开始使用磁芯、磁鼓等设备作为计算机的存储器。这个时期的计算机主要承担的是科学计算任务。

第二代计算机（1955—1960）使用晶体管。晶体管体积小、功耗低、速度快、可靠性高、成本低廉，适合大批量生产。晶体管是 1947 年发明的，但直到 1955 年才开始使用到计算机上。这个时期，计算机发展提速，开始用于科学计算之外的事务处理，同时开始使用电话线进行计算机的数据传输，这就是计算机网络的萌芽阶段。

20 世纪 60 年代开始使用集成电路（Integrated Circuits, IC）制造计算机，进入了第三代计算机时代（1963—1975）。集成电路是将大量的电子元器件，如电阻、电容、晶体管等制作在一个面积很小（以  $\text{cm}^2$  计）的半导体芯片上，以完成特定的电子线路功能。相对晶体管，集成电路在体积、重量、能耗等方面具有绝对的技术优势。在这个时期，另一个与计算机相关的重大事件是发射了同步通信卫星，由此实现了地面与空间的数据通信。

到了 70 年代，计算机核心的处理和电路可以被集成在单个芯片上，即计算机的 CPU（Central Processing Unit, 中央处理器），计算机进入了大规模集成电路的第四代（1975 年至今）。这个时期，最具影响力的是个人计算机的发展，也正是因为大规模集成电路的广泛应用而得以迅猛发展，使计算机走出科学家的实验室，摆到办公桌上，装入口袋里。

1965 年，戈登·摩尔（Gordon Moore）曾预言，集成电路的集成度（单位面积的元器件数量）每 18 个月提高 1 倍，价格下降一半。后来摩尔把时间修正到 2 年。这个预言被称为摩尔定律，此后的 40 多年的发展都验证了这个定律。

## 1.2.2 软件史

第一代计算机使用二进制代码编程，代码是计算机直接执行的指令代码。这个时期还没有软件的概念，只有编程（programming）。程序员需要非常熟悉机器指令代码，这个时期的程序员多为数学家和计算机专业工程师。机器代码编程极易出错，进而使用英文缩写表示机器代码，这就是汇编语言。汇编语言编写的程序最初要经过人工翻译成机器代码，后来翻译工作可以也通过程序来实现。

第二代计算机时期，计算机的硬件功能得到提升，对程序的要求自然随之提高。这个时期类似英文表达的高级编程语言被设计、开发出来。当时典型的高级编程语言有两个——FORTRAN 和 COBOL，前者主要用于科学计算，后者用于商业应用。这个时期，设计高级语言的是系统程序员，开发应用程序的人被称为应用程序员。这个时期另一个重要的变化是计算机界的巨头 IBM 公司放弃了软件随硬件捆绑的政策，第三方软件公司可以为用户提供编程服务，软件开发快速发展，有更多的软件公司进入计算机市场，形成了软件产业。

第三代计算机时期，有了操作系统（1.5 节将详细介绍）。这个时期出现了更多的高级编程语言和专门求解某一类问题的软件包，如统计软件 SPSS（Statistical Package for the Social Sciences，社会科学统计程序包）。同时，这个时期使用计算机的人也不再必须是计算机专业人员，因此计算机用户（user）这个重要的角色出现了。

第四代计算机时期，有了程序设计范式。结构化编程语言如 BASIC、C 语言等的出现，加快了软件开发速度，操作系统也开始朝着标准化的方向发展，各种应用软件（如文档处理、

电子表格、数据库系统等)大量出现,极大地推动着计算机应用的发展。尤其是 PC 的快速普及,带来的最明显变化是非专业人员成为计算机的主要用户。

20 世纪 90 年代出现了以图形界面为特征的操作系统,用户再也不需要记忆复杂的操作指令,通过鼠标对屏幕上的图形指针 (pointer) 点击操作就可以容易地使用计算机。计算机硬件的性能不断提升而价格越来越便宜 (摩尔定律),软件的功能越来越强,计算机已经“无所不能”。21 世纪初,软件产业已经超过了硬件的规模,同时出现了一个新的行业:服务。这个商业模式也是 IBM 首创的。

## 1.3 计算机模型

模型 (model) 是一种最常用的科学描述方法。模型是一种抽象表达,隐藏了复杂的细节,只展示其功能性部分。理解计算机系统的最好方法也是通过模型。广义的“计算机模型”是把计算机看成数据处理机。现代计算机模型定义了计算机的 5 个组成部分和程序存储原理。

### 1.3.1 数据处理机模型

计算机与 70 多年前相比有惊人的变化:今天的计算机不但体积更小,而且速度更快,价格也更便宜,但在逻辑上仍然很相似。因此,最早描述计算机功能的数据处理机模型到今天仍然是解释计算机原理最好的方法。数据处理机模型有多种,典型的是黑盒 (black box) 模型和具有程序能力的数据处理机模型。

#### 1. 黑盒模型

黑盒模型是用于许多学科的经典模型,计算机也用黑盒模型解释其基本原理。如图 1-3 所示,计算机就是一个黑盒子,它是数据处理的机器。

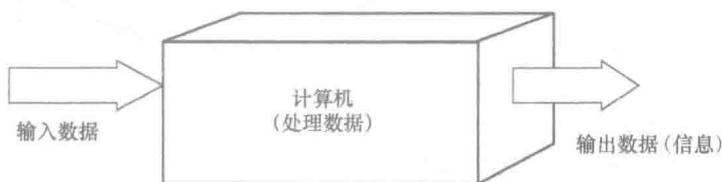


图 1-3 黑盒模型

计算机黑盒模型指出:数据输入到计算机中后,经过处理,输出结果。如果输入的数据相同,计算机的输出数据将能够重现;如果输入不同的数据,计算机的输出数据也随之改变。黑盒模型清晰地定义了计算机的功能:计算机是数据处理机。黑盒模型的缺点也很明显,它不能体现计算机的灵活性。

#### 2. 具有程序能力的数据处理机

具有程序能力的数据处理机如图 1-4 所示,它在黑盒模型的黑盒子上增加了一个程序 (program) 部分。程序在计算机中起到极为重要的作用,也是本书反复提及并从不同角度进行介绍的概念。程序种类繁多,有些程序的复杂度堪比目前已知的人类工程或系统,但可以简单地被理解为“按照预定的步骤进行工作”。

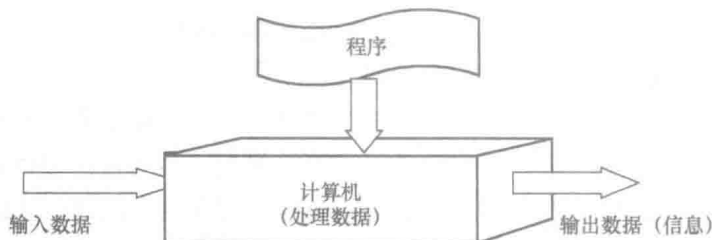


图 1-4 具有程序能力的计算机模型

在这个模型中，数据处理的结果取决于程序。因此，在同一个程序的控制下，相同的输入数据能够得到相同的输出数据。反之，如果程序不同，相同的输入数据也可能得到不同的输出数据。例如，程序控制计算机对一组数据分别进行累加和排序，那么得到的结果是完全不同的。进一步，不同的数据采用了不同的程序也可能产生相同的输出结果。

程序赋予了计算机的灵活性。计算机的强大处理能力体现为能够通过程序，让计算机完成不同的数据处理任务，得到期望的处理结果。

### 1.3.2 现代计算机模型

数据处理机模型有助于我们理解计算机的基本原理，能够进一步表述现代计算机模型，如图 1-5 所示。

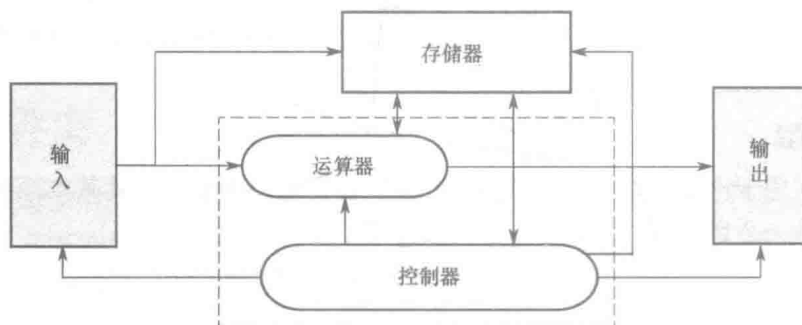


图 1-5 现代计算机模型

这个模型一直被称为冯·诺依曼模型，现在一些文献和教材还是沿用了“冯·诺依曼计算机”这一说法。有意思的是，大多数教科书中将其称为“计算机模型”（本书也是如此），但在冯·诺依曼的论文中，这个模型是“计算系统”。

计算机的诞生和发展得益于无数科学家们的才智，但回望它的历程，既有耀眼夺目的明星科学家们的卓越贡献，也有难以理清的争议。现代计算机究竟谁是第一贡献者<sup>①</sup>，就是其中的争议之一。

#### 1. 计算机的 5 个组成部分

现代计算机模型或冯·诺依曼模型定义了计算机的功能，还定义了计算机内部的结构。

<sup>①</sup> 在业界曾认为现代计算机模型的提出者是冯·诺依曼，但现在已经认定计算机的发明权属于阿塔纳索夫（John Atanasoff）和贝里（K·Berry），即 ABC 计算机（Atanasoff Berry Computer）的发明人。也曾一直认为“程序存储”的发明者是冯·诺依曼，但现在已经知道，最早提出这个概念的是宾西法尼亚大学 Moore 电子工程学院的 J·P·Eckert（第一台电子计算机 ENIAC 的发明者之一），而冯·诺依曼是首先于 1946 年公开发表了这个概念。