



信息通信技术普及丛书

走近 云计算

中国通信企业协会 组编
朱清峰 编著



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS



信息通信技术普及丛书

走近 云计算

中国通信企业协会 组编
朱清峰 编著

人民邮电出版社
北京

图书在版编目(CIP)数据

走近云计算 / 中国通信企业协会组编 ; 朱清峰编著

— 北京 : 人民邮电出版社, 2018. 12

(信息通信技术普及丛书)

ISBN 978-7-115-49475-7

I. ①走… II. ①中… ②朱… III. ①云计算—普及读物 IV. ①TP393.027-49

中国版本图书馆CIP数据核字(2018)第223108号

内 容 提 要

本书首先介绍了云计算的历史、概念和框架,然后从Web、虚拟化、分布式存储、分布式计算和安全5个维度全方位探讨了云计算的关键技术,接下来从应用模式、产品化、开源项目和标准化4个角度讲述了云计算的产业发展,并描绘了云计算在大数据、物联网、人工智能和虚拟现实等前沿领域的应用,最后结合阿里云给出了云计算在我们实际生活和工作中的应用体验。

本书适合计算机和通信行业相关技术人员以及对云计算技术感兴趣的读者阅读,也可以作为高等院校相关课程的参考书。

◆ 组 编 中国通信企业协会

编 著 朱清峰

责任编辑 李 静

责任印制 彭志环

◆ 人民邮电出版社出版发行 北京市丰台区

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

三河市祥达印刷包装有限公司印刷

◆ 开本: 700×1000 1/16

印张: 13.5

2018年12月第1版

字数: 250千字

2018年12月河北第1次印刷



定价: 69.00 元

读者服务热线: (010) 81055488 印装质量热线: (010) 81055316

防盗版热线: (010) 81055315



朱清峰

博士，高级工程师，中兴通讯股份有限公司软件资深专家。

1973年10月出生，1999年9月于东南大学仪器科学与工程系获得博士学位，同年10月进入中兴通讯股份有限公司，从事软件研发工作，拥有近20年通信产品软件研发经验。

《信息通信技术普及丛书》编委会

- 主 编:** 苗建华 中国通信企业协会会长
副 主 编: 刘桂清 中国电信集团有限公司副总经理
 顾晓敏 中国铁塔股份有限公司副总经理
 赵中新 中国通信企业协会副会长兼秘书长
 张同须 中国移动通信有限公司研究院院长
 张 涌 中国联通网络技术研究院院长
执行主编: 柏国林 中国通信企业协会副秘书长

编 辑 组

- 组 长:** 赵俊涅 中国通信企业协会综合业务发展部主任
副 组 长: 冯志宏 中国通信企业协会综合业务发展部副主任
 刘 婷 中国通信企业协会综合业务发展部副主任
 王建军 人民邮电出版社信通传媒图书出版中心主任

前言



在不少读者看来，计算机技术是一个推陈出新非常迅速的领域，云计算更是其中近年来非常热门的新技术。但实际上，这些新技术大多不是凭空出现的，它们均可以追溯到历史的故旧纸堆之中。

例如，作为云计算关键技术的虚拟化技术近年来大为热门，但实际上早在 20 世纪 60 年代，IBM 公司就已经在其大型机中成功应用了虚拟化技术。

又例如，Google 公司在 2004 年发表的关于 MapReduce 的论文一度使 MapReduce 成为云计算中最为引人注目的技术热点（最近几年来大家已经发现 MapReduce 并不是万能灵药），但实际上 MapReduce 的编程思想可以追溯到 20 世纪 60 年代约翰·麦卡锡发明的 Lisp 语言。

再例如，分布式系统的一致性问题是云计算的关键技术问题之一，Paxos 算法作为该问题最为经典的解决方案，是 Lamport 在 1990 年提出的。而 Paxos 也不是凭空出现的，它还可以继续追溯到 20 世纪 80 年代的拜占庭将军问题和 20 世纪 70 年代的两军问题。

因此，从这个意义上来说，笔者认为研究具体的代码实现固然重要，但理清其背后的设计思路也十分必要。目前市面上介绍云计算的书籍非常多，本书既不讲述编程工具的使用，也不讲述具体的软件代码，而是力求解释隐藏在这些工具和代码背后的原理和故事。

本书力求通俗易懂，适合在短时间内希望对云计算全貌有一个概括了解的读者。本书内容大致划分为 4 个组成部分。

第 1 部分包括第 1、2 章，主要介绍云计算的基本概念。第 1 章讲述了云计算的发展历史，第 2 章讲述了云计算的概念和框架。

第 2 部分包括第 3、4、5、6、7 章，主要介绍云计算的关键技术。第 3 章讲述了 Web 技术，第 4 章讲述了虚拟化技术，第 5 章讲述了分布式存储技术，第 6 章讲述了分布式计算技术，第 7 章讲述了安全技术。

第3部分包括第8、9章，主要介绍云计算的产业化和应用情况。第8章从应用模式、产品化、开源项目和标准化角度讲述了云计算的产业化情况，第9章讲述了云计算在大数据、物联网、人工智能和虚拟现实等前沿技术领域的应用情况。

第4部分是第10章，主要结合阿里云介绍云计算在我们实际生活工作中的应用体验。

最后，谨以此书献给妻子闵莉、儿子朱子涵、父亲朱安爱和母亲王云波，没有他们的支持和鼓励，就不会有此书的诞生。同时感谢人民邮电出版社的李静老师在本书编写出版过程中给予的指导和帮助。本书的不足之处欢迎各位读者不吝指正。

朱清峰

2018年4月于上海证大家园

目 录



第 1 章 云计算的历史	1
1.1 自动计算的思想史	2
1.2 计算机的历史	6
1.3 互联网的历史	9
1.4 网络计算的发展	12
1.5 云计算的兴起	15
第 2 章 什么是云计算	17
2.1 Google 数据中心之谜	18
2.2 Amazon 华丽的转身	21
2.3 Salesforce 的崛起	23
2.4 什么是云计算	26
2.5 云计算的框架	30
2.6 云计算的关键技术	32
第 3 章 Web	35
3.1 Web 的历史	36
3.1.1 WWW 的诞生	36
3.1.2 浏览器之争	37
3.1.3 Web 的标准	39
3.2 Web 技术体系	40
3.2.1 概述	40
3.2.2 Web 前端	41
3.2.3 Web 后端	43
3.2.4 设计演进	44

3.3	Web2.0	46
3.3.1	概述	46
3.3.2	Wiki	46
3.3.3	RSS/Atom	47
3.3.4	Blog	49
3.3.5	Podcast	50
3.3.6	Tag	50
3.3.7	SNS	51
3.4	云计算与 Web	52
3.5	移动互联时代的 Web	52
第 4 章	虚拟化	53
4.1	虚拟化的历史	54
4.2	存储虚拟化	55
4.2.1	RAID	56
4.2.2	NAS	57
4.2.3	SAN	58
4.3	网络虚拟化	59
4.3.1	VLAN	59
4.3.2	VPN	60
4.3.3	NFV	60
4.3.4	网络切片	61
4.4	软件虚拟化	62
4.4.1	指令级虚拟化	62
4.4.2	编程语言级虚拟化	63
4.4.3	运行库级虚拟化	64
4.5	计算机虚拟化	65
4.5.1	概述	65
4.5.2	CPU 的虚拟化	66
4.5.3	内存虚拟化	68
4.5.4	I/O 的虚拟化	69
4.6	云计算与虚拟化	70
4.6.1	虚拟机软件	70
4.6.2	弹性云服务	71
4.6.3	数据中心的虚拟化	72
4.7	容器技术	73

第 5 章 分布式存储	75
5.1 数据存储的新挑战	76
5.1.1 非结构化数据	76
5.1.2 大数据	77
5.1.3 计算机集群	77
5.2 分布式存储的历史	78
5.2.1 分布式文件系统的历史	78
5.2.2 数据库的历史	78
5.3 分布式系统的一致性	81
5.3.1 CAP 理论	81
5.3.2 ACID 与 BASE	82
5.3.3 两阶段提交	83
5.3.4 Paxos 与 Raft	84
5.4 分布式文件系统	86
5.4.1 GFS	86
5.4.2 HDFS	88
5.4.3 Ceph	89
5.4.4 混合存储	91
5.5 NoSQL 数据库	92
5.5.1 概述	92
5.5.2 键值存储	92
5.5.3 列族存储	93
5.5.4 文档存储	95
5.5.5 图存储	96
5.6 NewSQL 数据库	97
第 6 章 分布式计算	99
6.1 概述	100
6.1.1 Einstein@Home	100
6.1.2 蒙特卡罗方法	101
6.1.3 搜索引擎	102
6.1.4 什么是分布式计算	102
6.2 HPC	104
6.2.1 概述	104
6.2.2 异构加速	106
6.2.3 OpenMP	107

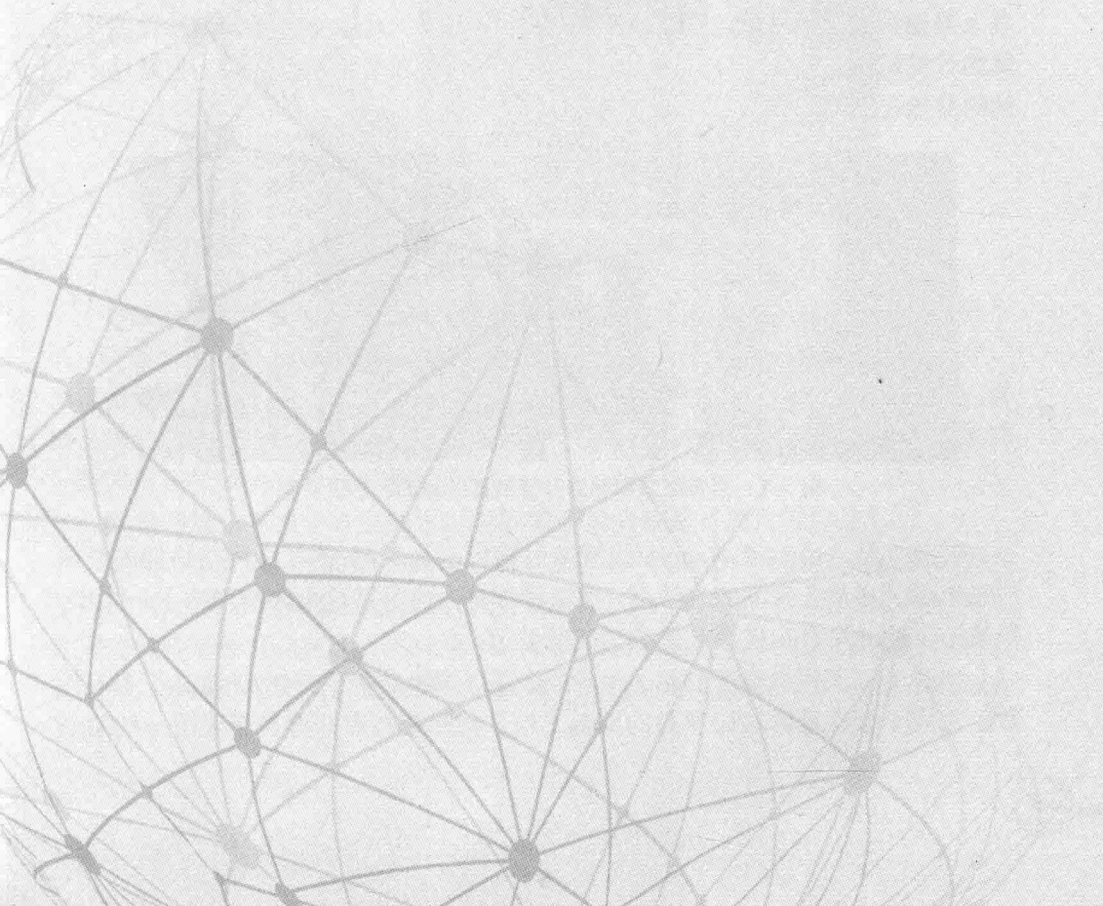
6.2.4	MPI	108
6.3	离线计算	109
6.3.1	概述	109
6.3.2	Map-Reduce	109
6.3.3	YARN	112
6.3.4	Mesos	114
6.3.5	Tez	115
6.4	实时计算	116
6.4.1	概述	116
6.4.2	Storm	117
6.4.3	Spark	117
6.4.4	Flink	120
6.5	分布式计算的发展	121
第7章	云计算的安全	123
7.1	概述	124
7.2	基础设施的安全	126
7.2.1	操作系统漏洞	126
7.2.2	防火墙	127
7.2.3	虚拟化的安全	129
7.3	数据的安全	130
7.3.1	数据加密	130
7.3.2	数据容灾	133
7.3.3	数据隐私	134
7.4	应用的安全	135
7.4.1	数字签名	135
7.4.2	PKI	136
7.4.3	身份认证	137
7.4.4	Sandbox	138
第8章	云计算的产业	139
8.1	概述	140
8.2	云计算的应用模式	141
8.2.1	公有云	141
8.2.2	私有云	142
8.2.3	混合云	142
8.3	云计算产品	143

8.3.1	概述	143
8.3.2	VMware	144
8.3.3	Amazon	145
8.3.4	微软	146
8.3.5	Salesforce	147
8.3.6	阿里云	148
8.4	云计算开源项目	150
8.4.1	概述	150
8.4.2	虚拟化与容器	151
8.4.3	IaaS	152
8.4.4	分布式存储	154
8.4.5	分布式计算	154
8.4.6	PaaS	154
8.5	云计算的标准化	156
8.5.1	概述	156
8.5.2	CSA	157
8.5.3	DMTF	158
8.5.4	NIST	159
8.5.5	OGF	162
8.5.6	SNIA	163
8.5.7	其他	164
第9章	云计算的应用	167
9.1	云计算与大数据	168
9.1.1	大数据的历史	168
9.1.2	什么是大数据	169
9.1.3	大数据的应用	171
9.1.4	大数据的技术	171
9.1.5	云计算与大数据	174
9.2	云计算与物联网	174
9.2.1	物联网的历史	174
9.2.2	什么是物联网	176
9.2.3	云计算与物联网	177
9.2.4	边缘计算	177
9.2.5	雾计算	178
9.3	云计算与人工智能	179
9.3.1	人工智能的历史	179

9.3.2	什么是人工智能	181
9.3.3	机器学习与深度学习	182
9.3.4	云计算与人工智能	184
9.4	云计算与虚拟现实	184
9.4.1	虚拟现实的历史	184
9.4.2	虚拟现实与增强现实	186
9.4.3	云计算与虚拟现实	187
第 10 章	云计算体验	189
10.1	概述	190
10.2	云服务器	190
10.2.1	了解 ECS	190
10.2.2	创建 ECS 实例	191
10.2.3	连接 ECS 实例	192
10.2.4	ECS 实例的管理	193
10.3	个人博客	193
10.3.1	更换 ECS 镜像	193
10.3.2	购买域名	194
10.3.3	安装 WordPress	194
10.3.4	开始博客之旅	196
10.4	个人网盘	197
10.4.1	准备工作	197
10.4.2	安装 NextCloud	197
10.4.3	玩转个人网盘	198
10.5	论坛网站	199
10.5.1	安装 WDCP	199
10.5.2	安装 Discuz	200
10.5.3	使用论坛	201
10.6	企业门户	202
10.6.1	准备工作	202
10.6.2	安装 PHPCMS	202
10.6.3	使用 CMS 网站	203
10.7	电商网站	203
10.7.1	准备工作	203
10.7.2	安装 ECShop	203
10.7.3	使用电商网站	204

第 1 章

云计算的历史



1.1 自动计算的思想史

计算机的历史可以划分为两部分：一部分是计算机工程技术的发展历史，例如电子管、晶体管、大规模集成电路等电子器件的发明；另一部分是自动化计算思想的发展历史，例如布尔代数、符号系统、图灵机、 λ 演算等思想的产生。

真正意义上的计算机是 20 世纪中叶才被人类发明出来的，而自动计算的思想却可以追溯到 17 世纪。正是由于这些思想先驱们的不断努力，人类首先在脑海中构建了计算机的设计蓝图，之后才在技术层面制造出了实际可应用的计算机。换句话说，计算机从 20 世纪中叶占地 170 平米的庞然大物（ENIAC），已经演进到可以装进裤兜的智能手机，虽然体积越来越小，功能越来越强，但是其背后隐藏的逻辑却始终如一。

17 世纪，法国数学家莱布尼茨在发现微积分，并发明了沿用至今的微积分符号之后，进一步提出了一个宏伟的设想：创造一套覆盖人类知识的纲要，采用符号系统对之进行准确描述，最后通过这些符号之间的演算就能够揭示所有人类知识之间的关联。虽然他在 1673 年制造了一台能够实现加减乘除四则运算的计算机器，如图 1-1 所示，并因此被选为伦敦皇家学会会员，但是这距离他的梦想还是遥不可及。

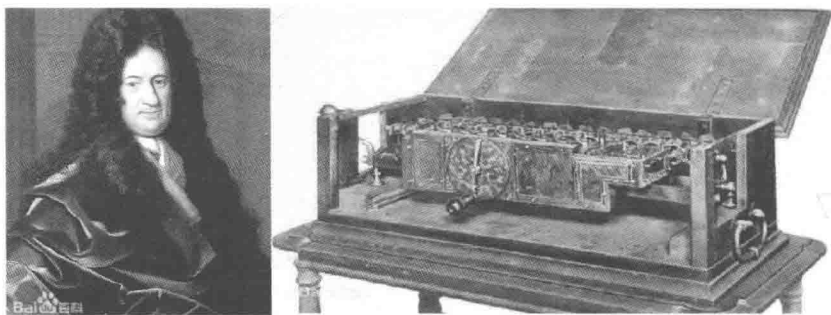


图 1-1 莱布尼兹和他的计算机器（来源：百度百科）

时至今日，虽然像 Google 的搜索引擎和 Facebook 的社交网络这样的系统已经承载了容量非常多的大数据，可以看作是朝着这个方向迈出了一大步，但显然还远远没有实现莱布尼兹当初的梦想。

1847 年，英国数学家乔治·布尔发表了《逻辑的数学分析》，建立了布尔代数，完成了数理逻辑这门学科的奠基工作。他把莱布尼兹的梦想采用一种简洁

的数学方式表现出来，采用符号表示逻辑概念，通过符号之间的代数运算规则来研究逻辑之间的关系。今天我们程序设计中采用的布尔值（TRUE 和 FALSE）就是为了纪念乔治·布尔而命名的。

1879年，德国数学家弗雷格出版了一本不到100页的小册子，名为《概念文字》。他认为布尔代数使用逻辑来推导逻辑的自循环方法不可接受，提出采用精确的语法和句法规则来构造形式化的人工语言。从这个角度来看，弗雷格提出的概念文字也可以看作是今天计算机程序设计语言的前身。

与此同时，德国数学家康托尔创立了集合论，首次将无穷这个概念引入数学领域。大名鼎鼎的德国数学家高斯曾经警告数学界：“我极力反对把无限当成一种完成的东西来使用，这在数学上是绝对不能允许的。无限只不过是言语上的一个比喻罢了。”

然而，康托尔还是将无穷以集合论的形式引入了数学。一方面，这促进了许多数学分支的产生，例如代数拓扑、群论、泛函分析；另一方面，这也像打开了潘多拉的盒子，给数学家们带来了无穷烦恼。如图1-2所示，以布尔、弗雷格和康托尔为代表的数学家完成了数理逻辑学的奠基工作。

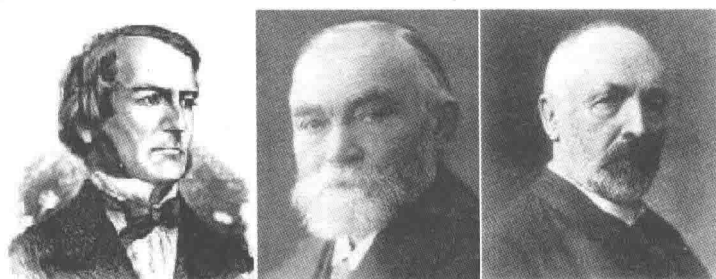


图1-2 数理逻辑学的奠基者：布尔、弗雷格和康托尔（来源：百度百科）

1902年，年轻的英国哲学家罗素给年逾50的弗雷格写了封信，提出了著名的“罗素悖论”。此时，正值弗雷格的新书《算数的基本法则》完稿付印之际。弗雷格看完来信后，在新书中补充到：“正当工作就要完成之时发现那大厦的基础已经动摇，对于一个科学工作者来说，没有什么比这更为不幸的了，罗素的一封来信正使我置身于这样的境地。”

罗素悖论中提出了许多例子，其中最著名的就是“理发师悖论”。

如图1-3所示，一位理发师的广告词如下：“我只为所有不给自己刮脸的人刮脸。”有一天，他从镜子里发现自己的胡子长了，请问他可以给自己刮脸吗？如果理发师不给自己刮脸，他就属于“不给自己刮脸的人”，那么他就应该给自己刮脸；如果理发师给自己刮脸，他就属于“给自己刮脸的人”，他就不应该给自己刮脸。

罗素悖论说明，集合论是自相矛盾的，数学变成了一门不严密科学，这引发了数学史上著名的第三次数学危机。经过众多数学家的不断努力，这次危机最终以公理集合论取代康托尔的朴素集合论而宣告解除。

解除了第三次数学危机以后，一直到1930年以前，整个数学界是非常乐观的，德国数学家希尔伯特提出了23个具有挑战性的数学命题。他认为整个数学的基础——数论和集合论已经公理化，如果能够继续证明这些公理系统是无矛盾的和完全的，那么整个数学的基础将非常牢靠。数学界一般也把这23个命题称为希尔伯特营救计划，如果这个计划成功的话，我们就可以为莱布尼茨的梦想找到一个坚实的数学基础。

1930年，24岁的哥德尔在一次数学基础研讨会上阐释了他在博士论文中的研究成果，即不完备性定理：即使把初等数论形式化之后，在这个形式的演绎系统中也可以找出一个合理的命题来，在该系统中既无法证明其为真，也无法证明其为假。当时与会的希尔伯特的学生冯·诺依曼立刻就意识到，这意味着希尔伯特的营救计划宣告失败。

哥德尔的不完备性定理从数学上证明了企图以形式化的方法一劳永逸地解决悖论问题的不可能性，同时也宣告试图通过数理逻辑学的方式实现莱布尼茨梦想彻底走入了死胡同。

与此同时，年轻的英国人图灵和他的美国导师丘奇无疑也在密切关注着数学界的这场大辩论，并各自进行了深入的思考。他们的难题有共同之处，都是探索关于计算的问题：如果我们有了无限计算能力的机器，哪些问题可以被解决？我们可以使它们自动地得以解决吗？是否还是有些问题无法解决？为什么？不同设计的各种机器是否具有相同的计算能力？

1936年，图灵在论文《论可计算数在判定问题中的应用》中提出，以布尔代数为基础，将逻辑中的任意命题可以用一种通用的机器来表示和完成，并能按照一定的规则推导出结论。这篇论文被誉为现代计算机原理的开山之作，它描述了一种假想的可实现通用计算的机器，后人称之为“图灵机”，图灵也因此被称为“计算机之父”。

如图1-4所示，所谓图灵机就是指一个抽象的机器，它有一条无限长的纸带，纸带分成了一个一个小方格，每个方格有不同的颜色。有一个机器头在纸带上移来移去。机器头有一组内部状态，还有一些固定的程序。在每个时刻，



图 1-3 罗素的理发师悖论