

高等教育规划教材



免费提供电子教案、
案例数据包及程序源代码

云计算与大数据 技术应用

安俊秀 靳宇倡 等编著



机械工业出版社
CHINA MACHINE PRESS

高等教育规划教材

云计算与大数据技术应用

安俊秀 靳宇倡 等编著



机械工业出版社

本书全面介绍了云计算与大数据的基础知识和主要技术。全书共 11 章, 主要内容包括云计算概述、大数据技术概述、虚拟化技术、数据中心、并行计算与集群技术、云存储技术、OpenStack、Hadoop、Spark、Storm 以及云计算仿真。本书注重实用, 实验丰富, 理论紧密联系实际, 使读者可以系统全面地了解云计算与大数据技术。

本书可作为高等院校云计算、大数据相关课程的教材, 也可以作为计算机相关专业的专业课或选修课教材, 同时也可以作为从事云计算与大数据技术相关的工作人员的参考用书。

本书配有教学资源(包括 PPT、案例数据包及程序源代码), 需要的老师可登录 www.cmpedu.com 免费注册, 审核通过后下载, 或联系编辑索取(QQ: 296698356, 电话: 010-88379739)。

图书在版编目(CIP)数据

云计算与大数据技术应用 / 安俊秀等编著. —北京: 机械工业出版社, 2019.5

高等教育规划教材

ISBN 978-7-111-63028-9

I. ①云… II. ①安… III. ①云计算—高等学校—教材 ②数据处理—高等学校—教材 IV. ①TP393.027 ②TP274

中国版本图书馆 CIP 数据核字(2019)第 124546 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

策划编辑: 王 斌 责任编辑: 王 斌

责任校对: 张艳霞 责任印制: 张 博

北京铭成印刷有限公司印刷

2019 年 7 月第 1 版·第 1 次印刷

184mm×260mm·13.75 印张·337 千字

0001—2500 册

标准书号: ISBN 978-7-111-63028-9

定价: 45.00 元

电话服务

客服电话: 010-88361066

010-88379833

010-68326294

封底无防伪标均为盗版

网络服务

机工官网: www.cmpbook.com

机工官博: weibo.com/cmp1952

金书网: www.golden-book.com

机工教育服务网: www.cmpedu.com

前 言

随着计算机技术的发展,数据量呈现爆炸式增长,传统 IT 技术已经无法满足业务的发展需求,因此云计算与大数据技术应运而生。目前,我国大部分的 Web 网站、移动应用、视频服务、游戏服务和电子商务平台都在应用云计算或大数据技术,并且云计算和大数据技术正迅速地在制造、金融、交通、医疗健康等各个行业得到广泛应用。随之而来的是广大企业对云计算与大数据人才的需求呈井喷式增长。为了满足社会发展的需求,云计算与大数据技术在高等院校计算机及相关专业的日常教学中逐渐占据重要地位,在计算机相关专业学生的培养计划中也逐步加入了云计算与大数据相关的课程。

本书将云计算与大数据技术整合进行讲解。首先介绍云计算技术,接着是对大数据技术进行概述,使读者对云计算与大数据的体系架构形成全局认识并了解二者之间的关系;在明晰基本理论的前提下,再对具体的虚拟化技术、数据中心、并行计算与集群技术、云存储技术、OpenStack、Hadoop 分布式大数据开发平台、Spark 大数据计算框架、Storm 基于拓扑的流数据实时计算框架、云计算仿真等内容进行具体介绍,并且穿插了相应的实践案例,使读者对云计算与大数据技术在具体实现上有更直观的认识,实现了理论与实践的有机结合。

本书非常适合开设云计算或者大数据技术相关课程的高校作为教材使用,也适合从事云计算与大数据技术相关工作的人员自用。在学习本书的内容之前,需要读者具备一定的计算机体系结构和计算机编程语言的基础。

本书共 11 章。

第 1 章是云计算概述,主要介绍云计算的定义、发展背景、基础架构和服务模式,以及云计算的部署模式、典型的云计算产品、云计算技术的新发展、我国的云计算产业现状。

第 2 章是大数据技术概述,包括大数据技术的产生、大数据的 4V 特征、大数据的主要应用及行业推动力量、大数据的关键技术、典型的大数据计算架构。

第 3 章对虚拟化技术做了较为详细的介绍,主要包括了虚拟化技术简介、虚拟化技术原理、常见的虚拟化技术解决方案、常见虚拟化技术的应用实践。

第 4 章介绍数据中心,主要包括数据中心的基本概念、数据中心的基本单元——服务器、数据中心的选址、数据中心的能耗。

第 5 章详细讲解了并行计算与集群技术,主要内容有并行计算概述、云计算基础架构——集群、并行计算的分类、并行计算相关技术、并行程序设计实践——MPI 编程。

第 6 章介绍了云存储技术,主要包括云存储概述、云存储与云计算、云存储的应用、云存储技术发展的关注点。

第 7 章介绍了 OpenStack 这一功能强大的 IaaS 平台,主要包括 OpenStack 架构和关键模块的介绍。

第 8 章介绍了 Hadoop 分布式大数据开发平台,包括 Hadoop 概述、分布式文件系统 HDFS、分布式计算框架 MapReduce、列式数据库 HBase 以及 Hadoop 开发环境的搭建。

第9章主要讲解了基于大规模数据实时处理的 Spark 内存计算框架，包括 Spark 概述、Spark 运行机制、Spark 运行模式、Spark RDD 和 Spark 的生态系统。

第10章详细讲解 Storm 这一基于拓扑的流数据实时计算框架，包括了 Storm 简介、Storm 原理及其体系结构、Storm-Yarn、搭建 Storm 开发环境及 Storm 应用实践。

第11章介绍云计算仿真，介绍了 CloudSim 云计算仿真系统、CloudSim 的模型使用场景、CloudSim 的应用实践。

本书由成都信息工程大学安俊秀教授和四川师范大学靳宇倡教授等共同编写。其中第1章、第2章、第5章、第9章由文仁强、安俊秀编写；第3章、第6章由陶武文、安俊秀编写；第4章由薛凯文、靳宇倡编写；第7章由安俊秀编写；第8章由刘明月、靳宇倡编写；第10章由王梓懿、靳宇倡编写；第11章由冯莉骄、靳宇倡编写。文仁强、薛凯文参与了本书的审阅工作。同时，本书的编写和出版还得到了国家自然科学基金项目（71673032）的支持。

尽管在本书的编写过程中，编者力求严谨、准确，但由于技术的发展日新月异，加之编者水平有限，书中难免存在错误和不足之处，敬请广大读者批评指正。如果有任何问题和建议，可发送电子邮件至 86631589@qq.com。

编者

目 录

前言

第 1 章 云计算概述	1
1.1 什么是云计算	1
1.1.1 云计算的定义	1
1.1.2 云计算的概念模型	2
1.1.3 云计算的特点	3
1.2 云计算技术发展背景	3
1.3 典型的云计算基础架构	7
1.4 云计算的主要服务模式	9
1.4.1 基础设施即服务 IaaS	9
1.4.2 平台即服务 PaaS	10
1.4.3 软件即服务 SaaS	11
1.4.4 三种服务模式之间的关系	12
1.5 云计算的主要部署模式	13
1.6 云计算是商业模式的创新	14
1.7 典型的云计算产品	15
1.7.1 Amazon 的 AWS	15
1.7.2 Windows Azure Platform	16
1.7.3 IBM 蓝云解决方案	17
1.7.4 阿里云	18
1.8 云计算技术的新发展	20
1.9 我国的云计算产业现状	23
1.9.1 政府推动云计算产业发展	23
1.9.2 我国云计算产业高速发展	24
习题	25
第 2 章 大数据技术概述	26
2.1 大数据技术的产生	26
2.1.1 大数据的基本概念	26
2.1.2 大数据产生的原因	26
2.1.3 大数据概念的提出	28
2.1.4 第四范式——大数据对科学研究产生的影响	30
2.1.5 云计算与大数据的关系	31
2.2 大数据的 4V 特征	31

2.3	大数据的主要应用及行业推动力量	32
2.3.1	大数据的主要应用	32
2.3.2	企业推动大数据行业发展	32
2.3.3	我国政府推动大数据行业发展	33
2.4	大数据的关键技术	34
2.5	典型的大数据计算架构	35
	习题	35
第3章	虚拟化技术	36
3.1	虚拟化技术简介	36
3.1.1	虚拟化技术的概念	36
3.1.2	虚拟化技术的分类	38
3.1.3	虚拟化技术的优势和劣势	42
3.1.4	虚拟化技术与云计算	43
3.2	虚拟化技术原理	43
3.2.1	虚拟机技术原理	43
3.2.2	CPU 虚拟化原理	44
3.2.3	内存虚拟化原理	46
3.2.4	网络虚拟化原理	47
3.3	常见的虚拟化技术解决方案	47
3.3.1	OpenStack	47
3.3.2	KVM	48
3.3.3	Hyper-V	49
3.3.4	VMware	49
3.3.5	Xen	50
3.3.6	Docker	51
3.4	常见虚拟化技术的应用实践	52
3.4.1	虚拟化环境的搭建	53
3.4.2	克隆虚拟机	54
3.4.3	虚拟机做快照	56
	习题	59
第4章	数据中心	60
4.1	数据中心的概念	60
4.1.1	数据中心的定义、作用及分类	60
4.1.2	数据中心的发展历程	61
4.1.3	数据中心的组成及建设原则	65
4.1.4	云计算、大数据时代的数据中心发展趋势	66
4.2	数据中心的基本单元——服务器	66
4.3	数据中心选址	69
4.4	数据中心的能耗	69

4.4.1	数据中心能耗评估	70
4.4.2	数据中心的主要节能措施	71
	习题	72
第 5 章	并行计算与集群技术	73
5.1	并行计算概述	73
5.1.1	并行计算的概念	73
5.1.2	并行计算的层次	75
5.1.3	并行计算机的发展	75
5.1.4	并行计算与分布式计算	77
5.1.5	并行计算与云计算	78
5.2	云计算基础架构——集群技术	79
5.2.1	集群的基本概念	79
5.2.2	集群系统的分类	80
5.2.3	集群文件系统	80
5.3	并行计算的分类	81
5.3.1	按 Flynn 分类	81
5.3.2	按应用的计算特征分类	82
5.3.3	按结构模型分类	83
5.4	并行计算相关技术	84
5.4.1	并行计算的关键技术	84
5.4.2	并行计算的性能估算	86
5.5	并行程序设计——MPI 编程	87
5.5.1	MPI 简介	87
5.5.2	一个简单的 MPI 程序实现	88
5.5.3	MPI 消息	90
5.5.4	MPI 的消息传递过程	90
5.5.5	MPI 常用基本函数	91
5.5.6	有消息传递的并行程序	91
	习题	93
第 6 章	云存储技术	94
6.1	云存储概述	94
6.1.1	云存储的概念	94
6.1.2	云存储系统的结构	95
6.1.3	云存储的实现基础	96
6.1.4	云存储的特性	98
6.2	云存储与云计算	99
6.3	云存储的应用	100
6.3.1	个人级云存储的应用	100
6.3.2	企业级云存储的应用	100

6.4	云存储发展的关注点	101
	习题	102
第7章	OpenStack——功能强大的 IaaS 平台	103
7.1	OpenStack 架构	103
7.2	计算服务模块 Nova	104
7.3	网络服务模块 Neutron	107
7.3.1	Neutron 的主要组件	107
7.3.2	Neutron 网络	109
7.4	块存储服务模块 Cinder	110
7.5	对象存储服务模块 Swift	111
7.6	身份认证模块 Keystone	115
7.7	镜像模块 Glance	119
7.8	仪表盘服务模块 Horizon	121
7.9	监控计量服务模块 Ceilometer	122
	习题	123
第8章	Hadoop——分布式大数据开发平台	124
8.1	Hadoop 简介	124
8.1.1	Hadoop 与分布式开发技术	124
8.1.2	Hadoop 的体系架构	125
8.1.3	Hadoop 集群的架构	127
8.2	分布式文件系统 HDFS	129
8.2.1	分布式文件系统概述	129
8.2.2	HDFS 的架构及读写流程	131
8.3	分布式计算框架 MapReduce	133
8.3.1	MapReduce 编程模型	133
8.3.2	MapReduce 数据流	135
8.3.3	MapReduce 任务运行流程	136
8.4	列式数据库 HBase	138
8.4.1	HBase 列数据库介绍	139
8.4.2	理解 HBase 的表结构	139
8.5	搭建 Hadoop 开发环境	142
8.5.1	相关准备工作	143
8.5.2	JDK 的安装配置	143
8.5.3	下载、解压 Hadoop 并配置 Hadoop 环境变量	144
8.5.4	修改 Hadoop 配置文件	145
8.5.5	将配置好的 Hadoop 文件复制到其他结点并格式化	146
8.5.6	启动、停止 Hadoop	146
8.5.7	运行测试程序 WordCount	147
	习题	148

第 9 章 Spark—基于内存的大数据计算框架	149
9.1 Spark 概述	149
9.2 Spark 的运行机制	150
9.3 Spark 的运行模式	152
9.3.1 Standalone 模式	153
9.3.2 Spark Yarn 模式	153
9.3.3 Spark Mesos 模式	157
9.4 Spark RDD	157
9.4.1 RDD 的特点	157
9.4.2 RDD 的创建	158
9.4.3 RDD 基本操作	160
9.4.4 RDD 持久化 (缓存)	162
9.4.5 Spark 共享变量	163
9.5 Spark 生态系统	165
9.5.1 Spark SQL	165
9.5.2 Spark Streaming	166
9.5.3 GraphX	169
9.5.4 MLlib	172
习题	173
第 10 章 Storm——基于拓扑的流数据实时计算框架	174
10.1 Storm 简介	174
10.2 Storm 原理及其体系架构	177
10.2.1 Storm 编程模型原理	177
10.2.2 Storm 体系架构	178
10.3 Storm-Yarn 简介	180
10.3.1 Storm-Yarn 的产生背景	180
10.3.2 Storm-Yarn 的体系架构	180
10.4 搭建 Storm 开发环境	181
10.4.1 Storm 安装说明	182
10.4.2 Storm 安装步骤	182
10.4.3 Storm 设置	185
10.4.4 Storm 的启动	186
10.4.5 Storm 的常用操作命令	186
10.5 Storm 应用实践	187
10.5.1 使用 Maven 管理 storm-starter	187
10.5.2 WordCountTopology 源代码分析	190
习题	193
第 11 章 云计算仿真	194
11.1 云计算仿真系统——CloudSim	194

11.1.1	CloudSim 基础	194
11.1.2	CloudSim 的体系结构	195
11.2	CloudSim 的模型使用场景	197
11.3	CloudSim 应用实践	198
11.3.1	准备环境	198
11.3.2	数据中心仿真实例	199
11.3.3	网络仿真实例	204
	习题	207
	参考文献	208

第1章 云计算概述

云计算(Cloud Computing)是基于互联网的相关服务的增加、使用和交付模式,通常涉及通过互联网来提供动态易扩展且常为虚拟化的资源,是并行计算(Parallel Computing)、分布式计算(Distributed Computing)和网格计算(Grid Computing)等的融合和发展,也是虚拟化(Virtualization)、效用计算(Utility Computing)、面向服务架构(SOA)等概念混合演进后商业实现的结果。本章将介绍什么是云计算、云计算的发展背景和特点、云计算的基础设施、掌握云计算的商业模式及云计算的主要服务模式。

1.1 什么是云计算

随着计算机技术的不断发展,云计算已经成为推动社会生产力变革的新生力量,那么什么是云计算?它有什么特点?以下一一进行介绍。

1.1.1 云计算的定义

到目前为止,业界对云计算的定义达100多种,尚没有一个统一的定义。云计算行业的领先者如Google、Microsoft等IT厂商及研究机构,依据各自的利益和各自不同的研究视角给出了以下对云计算的定义和理解。

1) 维基百科:云计算是一种动态扩展的计算模式,通过计算机网络将虚拟化的资源作为服务提供给用户;云计算通常包含基础设施即服务(Infrastructure as a Service, IaaS)、平台即服务(Platform as a Service, PaaS)、软件即服务(Software as a Service, SaaS)。

2) Google:将所有的计算和应用放置在“云”中,终端设备不需要安装任何软件,通过互联网来分享程序和服务。

3) 微软:云计算是“云+端”的计算,将计算资源分散分布,部分资源放在云上,部分资源放在用户终端,部分资源放在合作伙伴处,最终由用户选择合理的计算资源。

4) 互联网数据中心(International Data Corporation, IDC):云计算是一种新型的IT技术发展、部署及发布模式,能够通过互联网实时提供产品、服务和解决方案。

5) 美国国家标准与技术实验室(National Institute of Standards and Technology, NIST):云计算是一种无处不在的、便捷的、通过互联网访问的一个可定制的IT资源(IT资源包括网络、服务器、存储、应用软件和服务)共享池,是一种按使用量付费的模式。它能够通过最少量的管理或服务供应商的互动实现计算资源的迅速供给和释放。这是现阶段广为接受的云计算的定义。

6) 美国联邦云计算战略报告中,定义了四种云。

- 公有云:提供面向社会大众、公共群体的云计算服务。如Amazon云平台、Google App Engine等。公有云有很多优点,但最大的一个缺点是难以保证数据的私密性。

- 私有云：提供面向行业/组织内的云计算服务。如政府机关、企事业单位、学校等内部使用的云平台。私有云可较好地解决数据私密性问题，对数据私密性要求特别高的行业或组织，建设私有云将是一个必然的选择。例如 Window Azure 是私有云平台管理和软件。
- 社区云：提供面向社团组织内用户使用的云计算平台。如美国航天局（National Aeronautics and Space Administration, NASA）的 Nebula 云平台为 NASA 的研究人员提供快速的 IT 访问服务。
- 混合云：包含上述两种以上云计算类型的混合式云平台。

7) 我国对云计算定义：2012 年 3 月，在国务院政府工作报告中，将云计算作为国家战略性新兴产业，并给出了定义——云计算是基于互联网服务的增加、使用和交付模式，通常涉及通过互联网来提供动态易扩展且经常是虚拟化的资源。云计算是传统计算机和网络技术发展的产物，它意味着计算能力也可作为一种商品通过互联网进行流通。

云计算可以分为广义的和狭义的两类，狭义的云计算是指 IT 基础设施的交付和使用模式，即通过网络需求、扩展的方式获得所需的资源；广义的云计算是指服务的使用和交付模式，即通过网络按需求、扩展的方式获得对应的服务，它具有超大规模、虚拟化、可靠安全等特性。

简而言之，云计算是一种通过互联网以服务的方式提供动态可伸缩的虚拟化资源的计算模式。云计算的资源是分布式的，通过虚拟化技术动态易扩展。云计算是通过互联网提供的一种具有服务等级协议（Service-Level Agreement, SLA）的服务。SLA 协议是云服务提供商和客户之间的一份商业保障合同，而非一般的服务承诺。终端用户不需要了解“云”中基础设施的细节，不必具有相应的专业知识，也无须直接进行控制，只关注自己真正需要什么样的资源以及如何通过网络来得到相应的服务即可。

1.1.2 云计算的概念模型

从以上云计算的概念可知，云计算的实质是网络下的应用，其业务实现的概念模型，如图 1-1 所示。云计算包含了多层含义。

1) 用户的公共性。云计算面向各类用户，包括企业、政府部门、学术机构、个人等用户，也包括应用软件、中间件平台等“用户”。中间件是一种独立的系统软件或服务程序，分布式应用软件借助这种软件在不同的技术之间共享资源。中间件位于客户机/服务器的操作系统之上，管理计算机资源和网络通信，它是连接两个独立应用程序或独立系统的软件。

2) 设备的多样性。云计算用于提供服务的设备是多样的，既包括各种规模的服务器、主机、存储设备，也包括各种类型的终端设备，如计算机、智能手机、各种智能传感器、RFID（Radio Frequency Identification, 射频识别。它是一种通信技术，可通过无线电信号识别特定目标并读写相关数据，而无须在识别系统与特定目标之间建立机械或光学接触）设备等。

3) 商业模式的服务性。云计算的服务特性体现在两个方面：简化和标准的服务接口、按需计费的商业模式。

4) 提供方式的灵活性。云计算既可以作为一种公用设施，提供社会服务，即“公有云”，也可以作为企业信息化的集中计算平台来提供，即“私有云”。



图 1-1 云计算的概念模型

1.1.3 云计算的特点

与传统计算机系统相比，云计算具有以下特点。

1) 具有大规模并行计算能力。基于云端的强大而廉价的计算能力，为大粒度应用提供传统计算系统或用户终端所无法完成的计算服务。云计算系统的计算资源包括：CPU 运算资源、存储资源、网络带宽等。一般企业私有云有成百上千台服务器，有的甚至是上百万台服务器。

2) 资源虚拟化 (Virtualization) 和弹性调度。云计算系统的资源池中包括存储、处理器、内存、网络带宽等资源。它们以按需分配方式，为小粒度应用提供计算资源，实现资源共享。并且，云计算系统的规模可以动态地伸缩，满足不同的应用和不同的用户需求。同时，云计算系统中不同的物理机和虚拟机资源可根据客户需求动态分配。客户所获得的资源可能来自于北京的云计算资源，也可能来自于上海的云计算资源。虚拟化技术也是云计算的核心技术之一，包括了网络虚拟化、存储虚拟化、服务器虚拟化、操作系统虚拟化、应用虚拟化等。

3) 数据量巨大并且增速迅猛。由于在云计算环境下，人们既是信息的使用者，也是信息的创造者，导致互联网上的信息量剧增，那么如何使用这些数据为人们提供更好的服务成为目前的研究热点，并产生了典型的大数据处理技术，如 Hadoop、Spark、Storm 等，而这些技术都与云计算密不可分。

云计算还有一些其他特点：高可靠性，云计算应用了数据多副本容错，计算结点同构可互换等措施保证高可靠性；通用性，即同一个云可以支持不同的应用运行；高性价比，云计算相对低廉的价格使得用户使用起来更“实惠”。

1.2 云计算技术发展背景

从技术上来说，云计算的出现得益于网络带宽的飞速增长和互联网的飞速发展。网络带宽的迅速增长使得在互联网环境下进行快速数据处理成为可能。互联网提供了一种新的数据浏

览与交流的方式，用户既是信息的浏览者也是信息的创造者，从而产生了海量数据。这两个原因造成了使用智能终端和互联网的用户剧增，互联网成为用户日常生活所不可或缺的基础设施，得到了迅猛发展，促使了云计算的产生，并且已成为一种基础设施。

1. 飞速发展的网络技术

20 世纪 60 年代发生了第一波信息化革命，即计算机革命，很多传统企业紧跟这一轮信息化的浪潮，将计算机广泛应用到业务当中。20 世纪 90 年代发生了第二波信息化革命，即互联网革命。1987 年 9 月 14 日发出了中国第一封电子邮件：“Across the Great Wall we can reach every corner in the world. (越过长城，走向世界)”，揭开了中国人使用互联网的序幕，如图 1-2 所示。当时互联网的通信速率为 300bit/s。

```
DATE:      MON, 14 SEP 87 21:07 CHINA TIME
FROM:      "MAIL ADMINISTRATION FOR CHINA" <MAIL@ZE1>
TO:        ZORN@GERMANY, ROTERT@GERMANY, WACKER@GERMANY, FINKEN@UNIKAI
CC:        LHL@PARMESAN.WISC.EDU, FARBER@UDEL.EDU,

           JENNINGS%IRLEAN.BITNET@GERMANY, CIC%RELAY.CS.NET@GERMANY, WANG@ZE1,

           RZLI@ZE1
SUBJECT:   FIRST ELECTRONIC MAIL FROM CHINA TO GERMANY
```

"UEBER DIE GROSSE MAUER ERREICHEN WIE ALLE ECKEN DER WELT"

"ACROSS THE GREAT WALL WE CAN REACH EVERY CORNER IN THE WORLD"

DIES IST DIE ERSTE ELECTRONIC MAIL, DIE VON CHINA AUS UEBER RECHNERKOPPLUNG
IN DIE INTERNATIONALEN WISSENSCHAFTSNETZE GESCHICKT WIRD.

THIS IS THE FIRST ELECTRONIC MAIL SUPPOSED TO BE SENT FROM CHINA INTO THE
INTERNATIONAL SCIENTIFIC NETWORKS VIA COMPUTER INTERCONNECTION BETWEEN
BEIJING AND KARLSRUHE, WEST GERMANY (USING CSNET/PMDF BS2000 VERSION).

UNIVERSITY OF KARLSRUHE
-INFORMATIK RECHNERABTEILUNG-
(IRA)

INSTITUTE FOR COMPUTER APPLICATION OF
STATE COMMISSION OF MACHINE INDUSTRY
(ICA)

PROF. WERNER ZORN
MICHAEL FINKEN
STEFAN PAULISCH
MICHAEL ROTERT
GERHARD WACKER
HANS LACKNER

PROF. WANG YUN FENG
DR. LI CHENG CHIUNG
QIU LEI NAN
RUAN REN CHENG
WEI BAO XIAN
ZHU JIANG
ZHAO LI HUA

图 1-2 中国第一封电子邮件

1997 年 11 月，中国互联网络信息中心（China Internet Network Information Center, CNNIC）发布了第一次《中国互联网发展状况统计报告》：中国共有上网计算机 29.9 万台，上网用户数 62 万，CN 下注册的域名 4066 个，WWW 站点约 1500 个，国际出口带宽 25.408Mbit/s。1999 年 1 月国际出口带宽 143Mbit/s。截至 2018 年 6 月，我国国际出口带宽为 8826302Mbit/s。从中可以看到网络的发展速度之迅猛。2003 年，移动互联网开始兴起，随着 3G、4G 的出现，

移动互联网的网速实现了快速提升。如果说 3G、4G 充分连接了人与人，那么目前提出的 5G 则可以连接万物，尤其是在虚拟现实、自动驾驶、远程医疗等方面提供更丰富的体验。据中国互联网络信息中心（CNNIC）报告，截至 2018 年 12 月，中国网民规模达 8.29 亿，互联网普及率达到 59.6%。手机网民占比达 98.6%，手机支付习惯已经形成。手机网民规模及其占网民比例如图 1-3 所示。



图 1-3 手机网民规模及其占网民比例

云计算的核心其实是基于网络的应用，接入网络的带宽直接决定了企业使用云计算平台的质量。云计算服务使用方式简单、轻巧、方便，但是背后却消耗着大量的网络带宽，会大幅度增加互联网流量。云计算是一个可全球访问的资源结构，这意味着采用云计算结构的用户将是跨广域网的。

2. 万维网的发明与发展

1989 年，在欧洲粒子物理研究所工作的蒂姆·伯纳斯·李出于高能物理研究的需要发明了万维网（World Wide Web），也称为“Web”“WWW”“W3”，简称为 Web，如图 1-4 所示。4 年后，美国网景公司推出了万维网产品 Mosaic 浏览器，顿时风靡全世界。万维网的诞生给全球信息的交流和传播带来了革命性的变化，一举打开了人们通过互联网获取信息的方便之门。



<http://info.cern.ch>

伯纳斯·李：欧洲核子研究中心

世界上第一个网站

图 1-4 万维网的发明者及世界上第一个网站

万维网不等同于互联网，万维网只是互联网所能提供的服务之一，是靠着互联网运行的一项服务。也许大家想不到，尽管伯纳斯·李发明的万维网在 20 多年内创造出了无数的财富，但是创造者本人却坚持不对万维网申请专利，坚持让所有人都可以不付费而使用。由于他的杰出贡献，他被称为“互联网之父”。

Web 1.0 时代开始于 1994 年，其主要特征是大量使用静态的 HTML 网页来发布信息，并

开始使用浏览器来获取信息，这个时候主要是单向的信息传递。Web 1.0 的本质是聚合、联合、搜索，其聚合的对象是巨量、无序的互联网信息。Web 1.0 只解决了人对信息搜索、聚合的需求，而没有解决人与人之间沟通、互动和参与的需求。这个时期诞生了百度、Google、亚马逊等知名互联网企业。

在 21 世纪初期，正当互联网经济的泡沫破碎之际，Web 2.0（始于 2004 年 3 月）的兴起让网络迎来了一个新的发展高峰期。在 Web 2.0 中，软件被当成一种服务，万维网演化成了一个成熟的、为最终用户提供网络应用的服务平台，强调用户的参与、在线的网络协作、数据存储的网络化、社会关系网络、简易信息聚合（Really Simple Syndication, RSS）应用以及文件的共享。这个时候互联网上的信息传递变成了双向传递，用户既是信息的浏览者也是信息的创造者，Web 2.0 模式大大激发了创造和创新的积极性，使互联网变得生机勃勃。

在 Web 2.0 时代，Flickr、Myspace、Facebook、YouTube、Blog、Wiki 等社交网站的访问量已经远远超过传统门户网站。用户数量多以及用户参与程度高是这些网站的特点。因此，如何有效地为如此巨大的用户群体服务，让他们参与时能够享受方便、快捷的服务，成为这些网站不得不解决的一个问题。为了解决大型网站的访问量、并发量高、海量数据的问题，一般会考虑业务拆分和分布式部署，可以把那些关联不太大的业务独立出来，部署到不同的机器上，从而实现大规模的分布式系统。这也促进了云计算与大数据的产生。

3. 信息产业的发展演进

纵观整个信息技术的发展历史（如图 1-5 所示）看出，在不同时期，信息产业发展有两个重要的核心驱动力起着作用：硬件驱动力、网络驱动力。这两种驱动力量的对比和变化决定着产业中不同产品的出现时期以及不同形态的企业出现和消亡的时间。硬件驱动的时代诞生了 IBM、Microsoft、Intel 等企业。网络驱动的时代诞生了 Google、Yahoo、Amazon 等企业。

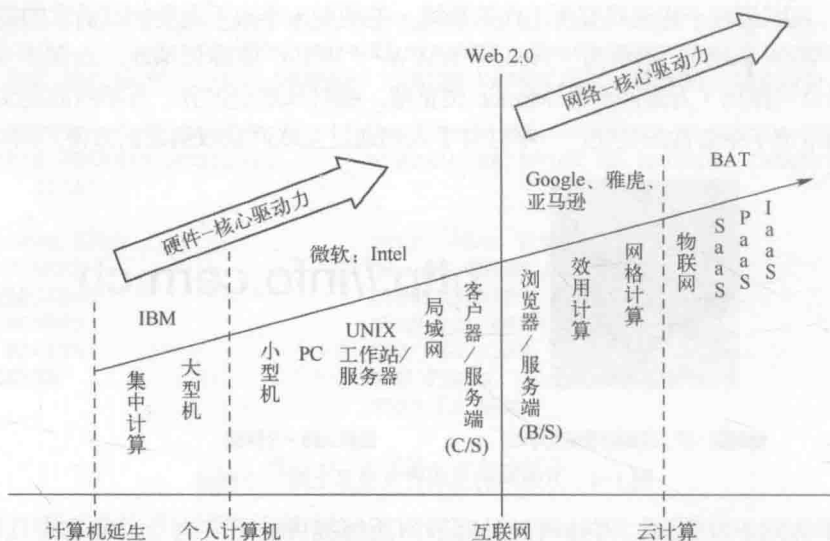


图 1-5 信息产业发展演进路线图

随着 2010 年掀起的第三波信息化革命，即移动互联网革命，世界正式进入大数据时代。与此同时，凭借 Google 文件系统搭建起来的 Google 服务器集群，为 Google 提供了强大的搜索速度与处理能力。于是，如何有效利用这些服务器资源，为更多的企业或个人提供强大的计