

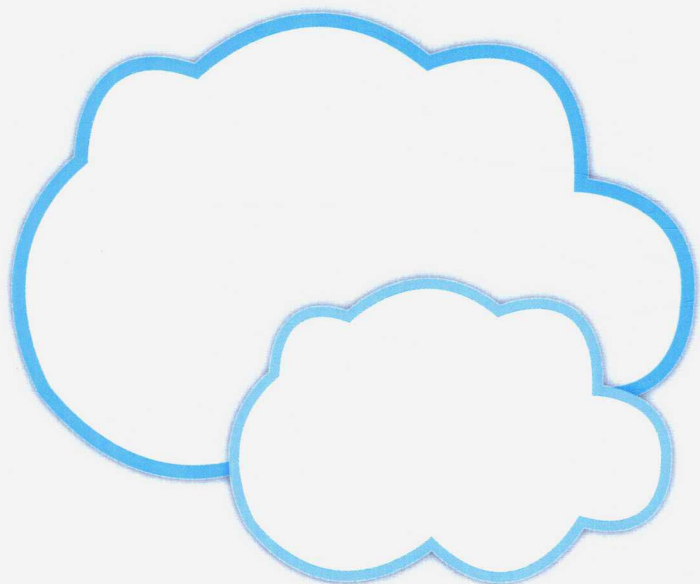
高等院校计算机专业教材

# 云计算

## 技术项目教程

秦 婷 张长华◎主编  
王伯爵 沈海龙 纪善国◎副主编

- ◎绪论
- ◎云计算的体系架构
- ◎云计算的关键技术
- ◎谷歌的云计算技术
- ◎开源云计算方案Hadoop
- ◎其他云计算方案
- ◎总结与展望
- ◎实验项目



知识产权出版社

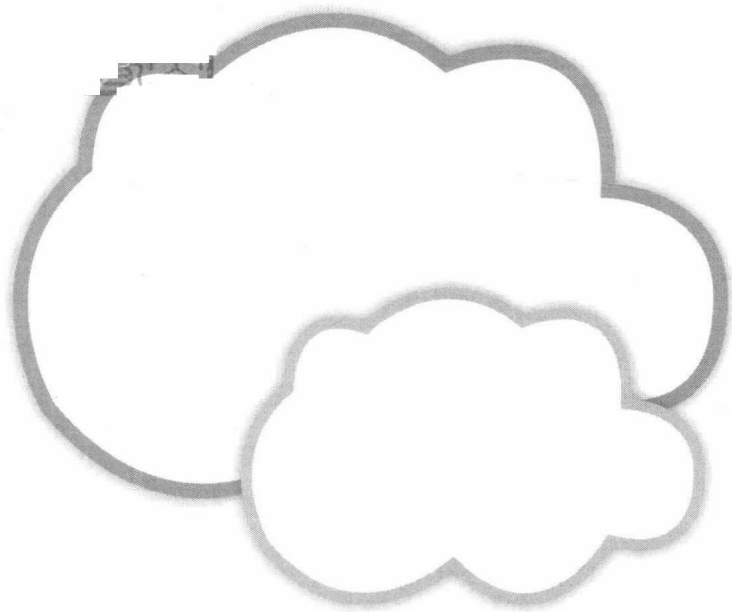
全国百佳图书出版单位

高等院校计算机专业教材

# 云计算 技术项目教程

秦 婷 张长华◎主编  
王伯爵 沈海龙 纪善国◎副主编

- ◎绪论
- ◎云计算的体系架构
- ◎云计算的关键技术
- ◎谷歌的云计算技术
- ◎开源云计算方案Hadoop
- ◎其他云计算方案
- ◎总结与展望
- ◎实验项目



知识产权出版社

全国百佳图书出版单位

## 图书在版编目 (CIP) 数据

云计算技术项目教程 / 秦婷, 张长华主编. — 北京: 知识产权出版社, 2016.12

ISBN 978-7-5130-3409-8

I. ①云… II. ①秦… ②张… III. ①计算机网络-教材 IV. ①TP393

中国版本图书馆 CIP 数据核字 (2016) 第 064068 号

## 内容提要

本书突出介绍了云计算技术的重要概念、关键技术和实践方法。第1章、第7章重点介绍云计算的概念、发展现状和趋势;第2章、第3章重点介绍云计算的体系架构、技术原理和一系列的关键技术;第4章、第5章和第6章分别介绍现今云计算平台的解决方案,包括谷歌的云计算技术方案和开源云计算项目 Hadoop 等;第8章安排了几个动手实践的实验项目,用以培养读者从事云计算开发的技术能力。

本书可作为全国高等学校计算机或信息等相关专业的学生教材,也可供云计算开发人员和爱好者学习和参考。

责任编辑: 徐家春

## 云计算技术项目教程

YUNJISUAN JISHU XIANGMU JIAOCHENG

秦 婷 张长华主编

出版发行: 知识产权出版社有限责任公司 网 址: <http://www.ipph.cn>

电 话: 010-82004826 <http://www.laichushu.com>

社 址: 北京市海淀区西外太平庄55号 邮 编: 100081

责编电话: 010-82000860转8573 责编邮箱: 823236309@qq.com

发行电话: 010-82000860转8101/8573 发行传真: 010-82000893/82003279

印 刷: 北京中献拓方科技发展有限公司 经 销: 各大网上书店、新华书店及相关专业书店

开 本: 787mm×1092mm 1/16 印 张: 14.25

版 次: 2016年12月第1版 印 次: 2016年12月第1次印刷

字 数: 300千字 定 价: 46.00元

ISBN 978-7-5130-3409-8

出版权专有 侵权必究

如有印装质量问题,本社负责调换。

# 前 言

随着通信技术和计算机技术的迅猛发展，通过网络访问远程的数据处理、存储和信息服务等计算服务的技术条件和需求也越来越成熟，于是“云计算”应运而生。近年来，随着云计算技术的迅猛发展，云计算得到了业界乃至全社会的广泛关注。云计算也被认为是继个人电脑革命、互联网革命以来，IT 产业的第三次变革。

云计算的服务设施不受用户端的影响，这意味着它们的规模和能力是不可限量的。当今，谷歌、亚马逊、微软和 IBM 等公司提供的云计算平台已经达到几十万乃至上百万台计算机的规模。由于规模经济的效应以及众多新技术的运用，加之拥有较高的资源利用率，云计算的性价比是传统的计算模式的 30 倍以上，这使得云计算成为一种划时代的技术。

当前，由于谷歌、亚马逊、微软和 IBM 等几家实力超群的大公司的大力推动，云计算发展极为迅速。一些小型的企业也开始跟随着云计算的浪潮受益，同时云计算也受到了各国政府以及广大用户的广泛关注。

谷歌、亚马逊、微软和 IBM 等大公司纷纷提出了自己的云计算平台，包括谷歌的 Google App Engine、亚马逊的 AWS、微软的 Windows Azure 等，他们采用了各不相同的设计架构和关键技术，本书将对之分别进行剖析。

中国政府近年来高度重视对云计算的发展。《中华人民共和国国民经济和社会发展第十二个五年规划纲要》和《国务院关于加快培育和发展战略性新兴产业的决定》(以下简称《决定》)均把云计算列为重点发展的战略性新兴产业。为了配合与落实国务院的《决定》，2010 年 10 月，中华人民共和国工业和信息化部与国家发展和改革委员会联合下发《关于做好云计算服务创新发展试点示范工作的通知》，确定北京、上海、杭州、深圳和无锡 5 城市先行开展云计算服务创新发展试点示范工作，并于 2011 年 10 月陆续下拨 6.6 亿的云计算专项扶植资金。

本书由内蒙古电子信息职业技术学院秦婷、河北建材职业技术学院张长华担任主编，天津国土资源和房屋职业学院的王伯爵、大连理工大学城市学院沈海

龙和山东省潍坊商业学校魏守峰担任副主编。

由于云计算技术发展迅速,加之作者水平有限、时间紧迫,书中难免存在疏漏和不当之处,恳请读者批评指正。

编者

2016年12月

# 目 录

第1章 绪论	1
1.1 云计算的概念	1
1.1.1 云计算的特征	2
1.1.2 云计算的优势	3
1.2 云计算的演进	3
1.2.1 云计算的由来	3
1.2.2 云计算的发展现状	5
1.3 云计算的应用场景	8
1.4 基于云计算的互联网应用	11
1.4.1 云办公	11
1.4.2 云存储	12
1.4.3 云输入法	13
1.4.4 云杀毒	14
1.5 云计算的发展环境	16
1.5.1 云计算与互联网	16
1.5.2 云计算与3G	18
1.5.3 云计算与物联网	19
1.6 本章小结	21
1.7 习题	21
第2章 云计算的体系架构	22
2.1 概述	22
2.2 云计算的体系结构	22
2.3 云计算的服务类型	25

2.3.1 IaaS	26
2.3.2 PaaS	27
2.3.3 SaaS	29
2.4 云计算的部署模式	31
2.4.1 公有云	31
2.4.2 私有云	33
2.4.3 混合云	34
2.4.4 行业云	35
2.5 本章小结	36
2.6 习题	36
<b>第3章 云计算的关键技术</b>	<b>37</b>
3.1 数据中心相关技术	37
3.1.1 数据中心网络设计	38
3.1.2 数据中心节能技术	39
3.2 虚拟化技术	40
3.2.1 什么是虚拟化	40
3.2.2 虚拟化技术的分类	41
3.2.3 云计算中的虚拟化技术	48
3.2.4 开源项目Xen	49
3.3 并行编程模型	50
3.3.1 并行编程	50
3.3.2 云计算的并行编程模型	52
3.4 数据存储技术	54
3.4.1 文件系统	54
3.4.2 分布式文件系统	55
3.4.3 GFS简介	58
3.5 数据管理技术	59
3.5.1 关系数据库	59
3.5.2 云计算对数据库技术的要求	61
3.5.3 NoSQL数据库	61
3.5.4 BigTable简介	62
3.6 资源管理与调度技术	63
3.6.1 副本管理技术	63
3.6.2 任务调度算法	63

3.6.3 任务容错机制 .....	65
3.7 QoS保证机制 .....	66
3.7.1 IaaS服务层的QoS保证机制 .....	66
3.7.2 PaaS和SaaS服务层的QoS保证机制 .....	67
3.8 安全与隐私保护 .....	68
3.9 本章小结 .....	70
3.10 习题 .....	70
<b>第4章 谷歌的云计算技术 .....</b>	<b>71</b>
4.1 概述 .....	71
4.2 谷歌云计算的体系架构 .....	71
4.2.1 设计思想 .....	72
4.2.2 整体架构 .....	74
4.3 GFS .....	77
4.3.1 系统结构 .....	78
4.3.2 容错机制 .....	80
4.3.3 系统管理技术 .....	81
4.4 MapReduce .....	82
4.4.1 产生背景 .....	82
4.4.2 编程模型 .....	83
4.4.3 执行流程 .....	84
4.4.4 单词计数问题 .....	85
4.5 Chubby .....	87
4.5.1 产生背景 .....	88
4.5.2 系统架构 .....	89
4.5.3 通信协议 .....	90
4.5.4 正确性与性能 .....	92
4.6 BigTable .....	93
4.6.1 产生背景 .....	94
4.6.2 数据模型 .....	94
4.6.3 系统架构 .....	96
4.6.4 主服务器 .....	96
4.6.5 子表服务器 .....	98
4.6.6 性能优化 .....	101



4.7 其他技术	103
4.7.1 Megastore	103
4.7.2 Dapper	103
4.8 谷歌的云计算产品	103
4.8.1 SaaS产品	103
4.8.2 PaaS	105
4.8.3 IaaS	105
4.8.4 云客户端	105
4.9 Google App Engine	108
4.9.1 Google App Engine简介	108
4.9.2 Google App Engine的架构	113
4.9.3 Google App Engine的使用	115
4.10 本章小结	121
4.11 习题	122
<b>第5章 开源云计算方案Hadoop</b>	<b>123</b>
5.1 Hadoop简介	123
5.1.1 Hadoop的产生	123
5.1.2 Hadoop的组成	123
5.1.3 Hadoop的架构	124
5.1.4 Hadoop的优势	126
5.2 HDFS	126
5.2.1 HDFS的体系结构	127
5.2.2 Hadoop的文件操作	128
5.2.3 Hadoop的容错支持	130
5.2.4 HDFS的访问接口	131
5.3 MapReduce	132
5.4 HBase	134
5.4.1 HBase的数据模型	134
5.4.2 HBase的系统架构	135
5.5 Hadoop vs Google云	137
5.5.1 HDFS vs GFS	137
5.5.2 HBase vs BigTable	138
5.6 本章小结	139
5.7 习题	139

第6章 其他云计算方案 .....	140
6.1 Amazon的云计算 .....	140
6.1.1 概述 .....	140
6.1.2 Amazon的云计算平台AWS .....	141
6.1.3 弹性计算云EC2 .....	142
6.1.4 简单存储服务S3 .....	145
6.1.5 简单队列服务SQS .....	146
6.1.6 简单数据库服务SDB .....	147
6.1.7 AWS的综合使用举例 .....	149
6.2 微软的云计算 .....	149
6.2.1 概述 .....	149
6.2.2 Live和 Online解决方案 (SaaS) .....	150
6.2.3 Windows Azure Platform解决方案 (PaaS) .....	151
6.2.4 动态云解决方案 (IaaS) .....	155
6.3 VMware的云计算 .....	156
6.3.1 概述 .....	156
6.3.2 云基础架构及管理 (IaaS) .....	157
6.3.3 云应用平台 (PaaS) .....	159
6.3.4 终端用户计算解决方案: 桌面虚拟化产品 (SaaS) .....	159
6.4 IBM的云计算 .....	161
6.4.1 概述 .....	161
6.4.2 “蓝云”解决方案的“6+1”场景 .....	162
6.4.3 “蓝云”解决方案的优势 .....	164
6.5 本章小结 .....	165
6.6 习题 .....	165
第7章 总结与展望 .....	166
7.1 云计算的行业应用 .....	166
7.1.1 电信行业 .....	166
7.1.2 金融行业 .....	168
7.2 云计算产业分析 .....	170
7.2.1 云计算产业体系 .....	170
7.2.2 全球云计算产业发展现状 .....	171
7.2.3 我国云计算产业发展现状 .....	172

7.3 各国政府的云计算发展战略	173
7.3.1 外国政府的云计算行动	173
7.3.2 我国政府的云计算行动	174
7.4 我国云计算发展面临的机遇和挑战	176
7.4.1 我国云计算发展面临的机遇	176
7.4.2 我国云计算发展面临的挑战	176
7.4.3 我国云计算未来发展思考	177
7.5 云计算的标准化进展	177
7.6 云计算面临的安全问题	178
7.6.1 云计算安全	178
7.6.2 云计算的法律环境	179
7.7 云计算的未来	179
7.8 本章小结	180
7.9 习题	180
<b>第8章 实验项目</b>	<b>181</b>
8.1 Virtual PC的安装、配置和使用	181
8.1.1 实验目的	181
8.1.2 实验内容	181
8.2 Xen的安装、部署和配置	181
8.2.2 实验要求	181
8.2.3 实验内容	182
8.2.4 实验步骤指导	182
8.3 Hadoop和HBase的安装、部署和配置	186
8.3.2 实验内容	187
8.3.3 实验要求	187
8.3.4 实验步骤指导	187
8.4 MapReduce编程	196
8.4.1 实验目的	196
8.4.2 实验内容	196
8.4.3 实验要求	196
8.4.4 实验步骤指导	196
8.4.5 实训作业	205

8.5 SaaS应用编程 .....	207
8.5.2 实验内容 .....	207
8.5.3 实验要求 .....	207
8.5.4 实验步骤指导 .....	207
8.5.5 实训作业 .....	208
8.6 Nutch安装、部署和主题搜索 .....	208
8.6.1 实验目的 .....	208
8.6.2 实验内容 .....	208
8.6.3 实验要求 .....	208
8.6.4 实验步骤 .....	209
8.7 本章小结 .....	211
参考文献 .....	212

# 第1章 绪论

## 1.1 云计算的概念

近年来,随着云计算技术的迅猛发展,云计算得到了业界乃至全社会的广泛关注。云计算也被认为是继个人电脑革命、互联网革命以来,IT产业的第三次变革。

美国国家标准与技术研究院(NIST)是这样定义云计算的:云计算是对基于网络的、可配置的共享计算资源池能够方便地、按需地访问的一种模式。所谓的共享计算资源池包括网络、服务器、存储、应用和服务。这个共享计算资源池就是我们所说的“云”。

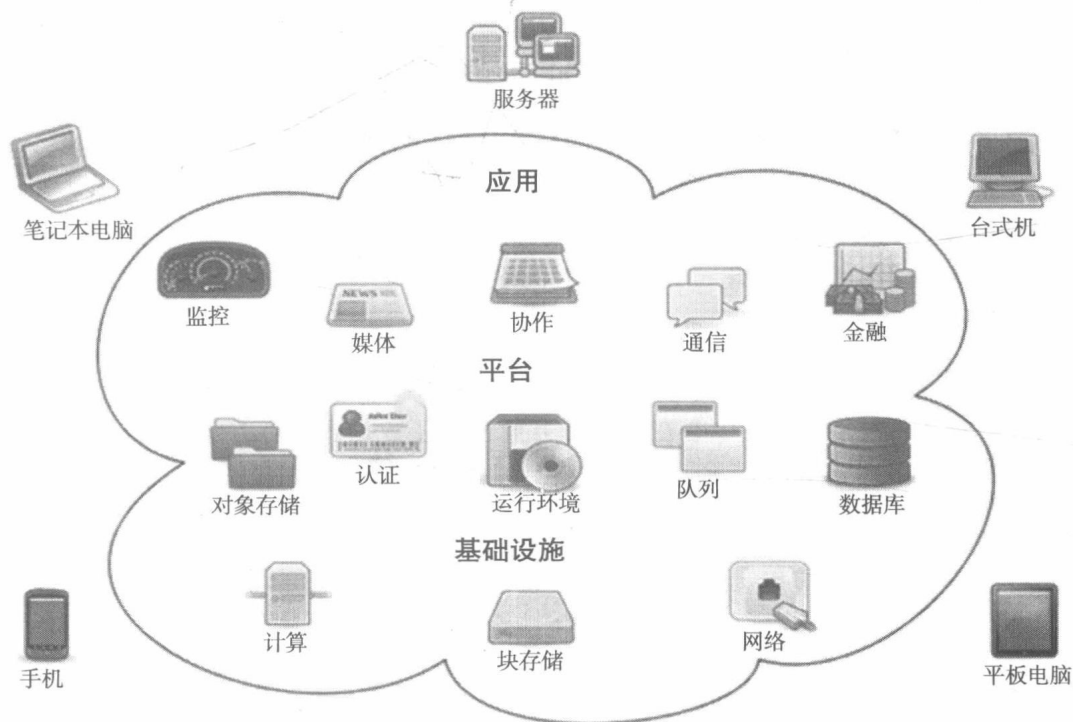


图1-1 云计算

关于云的定义，笔者认为主要有两个关键的方面：一方面，“云”能够自动化地、可靠地整合和管理各种地理上分散的或异构的计算资源，形成一朵由不同计算资源融合成的“云”；另一方面，“云”能够按需地、弹性地、可计量地，以计算能力、存储能力或软件服务的形式，向用户提供整合好的计算资源，使这些分散的资源对用户透明，也像是一朵“云”在为会员服务（图 1-1）。

所以，云计算与传统的计算资源的获取方式相比，就好比供电从原始的单台发电机模式转向了电厂集中供电的模式。它意味着计算能力、存储能力和软件服务能力可以作为一种商品进行流通，就像煤气、水电一样，取用方便，费用低廉。不同在于，它是通过互联网进行传输的。

技术上，云计算是并行计算、分布式计算和网格计算的应用和发展，或者说是这些计算机科学概念的一种商业实现。云计算是虚拟化、效用计算、IaaS（Infrastructure as a Service，基础设施即服务）、PaaS（Platform as a Service，平台即服务）、SaaS（Software as a Service，软件即服务）等概念混合演进并跃升的结果。所以，与其说云计算是一种新技术，不如说它是一种新的技术整合模式，或是新的商业概念。

### 1.1.1 云计算的特征

云计算主要具有以下特征。

（1）超大规模：大多数的云计算中心都有相当的规模，如谷歌云计算中心已经拥有几百万台服务器，而亚马逊、IBM、微软、雅虎等企业所掌握的云计算规模也毫不逊色。同时，云计算中心能够通过整合和管理这些数目庞大的计算机集群，来赋予用户前所未有的计算能力和存储能力。

（2）虚拟化：即抽象化。云计算支持用户在任意位置使用各种终端获取应用服务。所请求的资源来自“云”，而不是固定的有形的实体。应用在“云”中某处运行，但实际上用户无需了解，也不用担心应用运行的具体位置，即对用户“透明”。

（3）高可靠性：云计算中心在软硬件方面使用了数据多副本容错、计算节点同构可互换等措施来保障服务的高可靠性，使用云计算比使用本地计算机更可靠。

（4）通用性：云计算中心不是为了某些特定的应用而存在，在“云”的支撑下可以构造出千变万化的应用，同一个“云”可以同时支撑不同的应用运行，并保证其服务质量。

（5）高可扩展性：即高弹性。“云”的规模可以动态伸缩，以满足应用和用户规模增长的需要。

（6）按需服务：“云”是一个庞大的资源池，用户可以按需购买，就像自来水、电、煤气等根据用户的使用量来计费。

(7) 廉价: 由于“云”的特殊容错措施可以采用极其廉价的节点来构成云,“云”的自动化集中式管理使大量企业无需负担日益高昂的数据中心管理成本,“云”的通用性使资源的利用率较传统系统大幅提升,因此用户可以充分享受“云”的低成本优势,获得强大的计算能力和存储能力,甚至只要花费几百美元、几天时间就能完成以前需要数万美元、数月时间才能完成的任务。

(8) 节能环保: 云计算技术可以将许多地理上分散的低端机器的工作能力整合到一起,来提升资源的使用效率,同时一般由专业的管理团队运维,所以其电源使用效率比普通企业的数据中心出色很多。

这些特点使得云计算具有前所未有的优势,获得了强劲的发展和业界广泛的关注。

### 1.1.2 云计算的优势

从用户体验的角度,对个人用户而言,在云计算的背景下会出现各种丰富多彩的基于互联网的服务,这些服务功能强大,可以随时随地接入,无需客户端,只需要浏览器就能够轻松地访问,也无需为软件的升级和病毒的感染而操心;对企业用户而言,可以按需地、廉价地获得计算能力和存储能力,即将企业的 IT 服务迁移到“云”上,这样企业的管理更加方便,使其有利于专注于其自身领域的业务。

从成本的角度,对个人用户而言,他们所需的服务运行在“云”上,所以他们只需要一个简单的可以接入网络的终端即可,如手机、平板电脑等;对企业用户而言,可以利用先进的云技术来降低企业初期的投资成本和后期的维护成本,还可以将 IT 服务迁移到第三方提供的“云”上,从而降低 IT 部门的运维成本,节省企业的管理资源。

## 1.2 云计算的演进

### 1.2.1 云计算的由来

和许多其他的伟大发明和技术一样,云计算看似是人类灵光一现的产物,但真正探究、溯源起来,却是通过一代人甚至几代人的积累演变而成的。云计算的发明是整个 IT 产业自然发展和演化的必然结果,这些演化既表现在思想上,又涉及技术方面的发展与演进。

“云计算”的名字最早是由谷歌的 CEO (首席执行官) 埃里克·施密特在担任 Sun 公司 CTO (首席技术官) 时偶然想到的,其在思想方面的发展经历了电厂模式、效用计算、网格计算和云计算 4 个主要的阶段,才发展到现在比较成熟的水平 (图 1-2)。

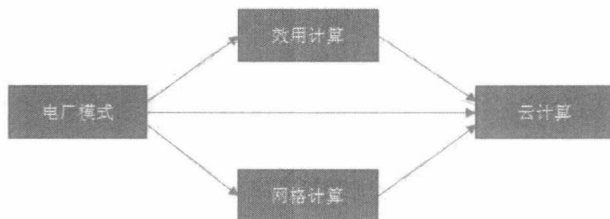


图1-2 云计算的思想演进

电厂模式的实质就是规模经济的效应，即形成规模之后可以有效降低成本。还是源于我们前面举到的电厂和发电机的例子，利用电厂的规模效应可以降低成本，从而降低价格，同时用户使用起来更加方便，可以按需使用、按量付费，不用考虑维护和购买任何发电设备。

人工智能之父麦肯锡在 1961 年提出了“效用计算”（Utility Computing）的概念，它是对电厂模式的借鉴和进一步发展，其目标是整合分散在各地的服务器、存储系统和应用程序，共享给多个用户，让用户能够像把灯泡插入灯座一样来使用计算机资源，并且根据其所使用的量来付费。

网格计算（Grid Computing）研究如何把一个需要十分巨大的计算能力才能解决的问题分成许多小的部分，然后把这些小的部分分配给大量性能较低的计算机来处理，最后把这些计算结果综合起来攻克大问题。由此可见，网格计算的实质与效用计算十分接近，但是侧重点略有不同。

云计算的核心思想源于人们对电厂模式的憧憬，又和效用计算、网格计算十分类似，即希望 IT 资源和服务能够像电力那样成本低廉、使用方便。与早期的思想不同的是，在 IT 技术高速发展的今天，无论是技术，还是用户的需求，都已经逐渐成熟并具有了一定的规模，所以云计算已经有了坚实的发展基础。

除了思想上的演化，还有这样几个重要的时间点，极大地推动了云计算的诞生、发展和成熟，也可以看作是云计算发展的几个重要的契机。

1999 年，当时甲骨文的高管 Marc Benioff 认为 Web 应用取代桌面应用是未来的趋势，于是创建了 Salesforce 这家以销售在线 CRM（Customer Relationship Management，客户关系管理）系统为主的互联网公司，同时首先定义了 SaaS（Software as a Service，软件即服务）的概念。SaaS 的实质就是让软件以在线服务的方式提供给用户，同时免去用户安装和更新等繁琐的步骤。从此以后，SaaS 受到了业界的广泛关注并且得到了良好的发展，这无疑为云计算的诞生提供了技术方面和商业运作方面的基础。

2003 年，在操作系统原理会议（SOSP）上，谷歌发表了有关 GFS（Google File System，谷歌文件系统）分布式存储系统的论文；2004 年，在 OSDI 大会上，谷歌发



表了有关 MapReduce 分布式处理技术的论文；2006 年，在 OSDI 大会上，谷歌发表了关于 BigTable 分布式数据库的论文。这三篇重量级论文的发表及其相关的开源技术极大地普及了云计算中非常核心的分布式技术。

2006 年初，亚马逊推出了 AWS（Amazon Web Service）的第一款产品 S3（Simple Storage Service，简单存储服务）云存储服务；2006 年 8 月，亚马逊推出了另一款 AWS 产品 EC2（Elastic Computing Cloud，弹性计算云）云基础设施服务。AWS 就是基于亚马逊将计算资源提供给用户的思想，S3 就是将服务器端的海量存储资源提供给用户使用，EC2 则是将整合好的计算资源作为一种服务或是商品提供给用户。AWS 相关服务的推出，标志着云计算商业产品的出现，所以亚马逊也被业界认为是云计算的先行者。

### 1.2.2 云计算的发展现状

云计算作为当前被炒得火热的新概念，由于有亚马逊、谷歌、IBM 和微软等几家实力超群的大公司的大力推动，发展极为迅速。一些小型的企业也开始跟随着云计算的浪潮受益，同时云计算也受到了各国政府及广大用户的广泛关注（图 1-3）。



图 1-3 云计算受到了业界的广泛关注和追捧

亚马逊的简单存储服务（S3）和弹性计算云（EC2）为企业提供存储和计算服务，收费的服务项目包括存储服务器、带宽、中央处理器（CPU）资源及月租费。月租费与电话月租费类似，存储服务器、带宽按容量收费，CPU 根据时长（小时）运算量收费。亚马逊把云计算做成一个大生意并没有花费太长的时间——不到两年时间，亚马逊上的注册开发人员达 44 万人，还有为数众多的企业级用户。由第三方统计机构提供的数据