

Cloud Computing Practice

由入门到实践，对云计算技术进行全面、  
深入的剖析，案例典型详实，  
具有很高的实用价值。

张德丰◎主编

# 云计算 实战

Cloud Computing  
Practice

- 合理、完善的体系结构
- 原理与技术的完美结合
- 云计算应用的最佳手册
- 教学与科研的最新成果

清华大学出版社

Cloud Computing Practice

张德丰◎主编

# 云计算实战

Cloud Computing  
Practice

清华大学出版社

北京

## 内 容 简 介

云计算是继 20 世纪 80 年代大型计算机到客户端-服务器的大转变之后的又一巨变,它描述了一种基于互联网的相关服务的增加、使用和交付模式,通常通过互联网来提供动态、易扩展而且经常是虚拟化的资源,广泛应用于云物联、云安全、云存储。本书介绍了云计算的基本概念、原理及实际应用,重点阐述了云计算领域具有代表性的 Google、Amazon、微软和 VMware 四家公司的云计算核心原理、使用方法及应用。主要内容包括 Google 云计算实现、Hadoop 分布式文件系统、Amazon 云计算、VMware 云计算、Windows Azure 云计算及云计算其他应用软件等。

本书结合具体实例来讲解相关概念及原理,实用性较强,适合作为大学高年级和研究生云计算课程教材,也可作为云计算研究开发人员、爱好者的参考用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。  
版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

### 图书在版编目(CIP)数据

云计算实战/张德丰主编. —北京:清华大学出版社,2012.6

ISBN 978-7-302-28931-9

I. ①云… II. ①张… III. ①计算机网络 IV. ①TP393

中国版本图书馆 CIP 数据核字(2012)第 111018 号

责任编辑:钟志芳

封面设计:刘超

版式设计:文森时代

责任校对:王云

责任印制:

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

印 刷 者:

装 订 者:

经 销:全国新华书店

开 本:185mm×260mm 印 张:25 字 数:573 千字

版 次:2012 年 6 月第 1 版 印 次:2012 年 6 月第 1 次印刷

印 数:1~4000

定 价:48.00 元

---

产品编号:047252-01

# 前 言

随着网络带宽的不断增长，通过网络访问非本地的计算服务条件（包括数据处理、存储和信息服务等）越来越成熟，于是就有了今天被称作“云计算”（Cloud Computing）的技术。云计算在 2008 年成为一个热门话题，到 2009 年时云计算的概念已十分普及。之所以称作“云”，是因为计算设施不在本地而在网络中，用户不需要关心它们所处的具体位置，于是就像以前画网络那样用“一朵云”来代替了。

云计算并不是一个年度忽悠词，而是与从大型机到 PC、PC 到互联网一样，代表了一次深刻的变革，开启了终端+“云计算”的大数据时代的巨幕。

云计算是基于互联网的相关服务的增加、使用和交付模式，通常通过互联网来提供动态、易扩展且经常是虚拟化的资源。云是网络、互联网的一种比喻说法。过去在图中往往用云来表示电信网，后来也用来表示互联网和底层基础设施的抽象。狭义的云计算指 IT 基础设施的交付和使用模式，指通过网络以按需、易扩展的方式获得所需资源；广义的云计算指服务的交付和使用模式，指通过网络以按需、易扩展的方式获得所需服务。这种服务可以是 IT 和软件、互联网相关，也可以是其他服务，这也意味着计算能力可以作为一种商品通过互联网进行流通。

云计算与当今同样备受关注的 3G 和物联网是什么关系呢？是互为支撑、交互辉映的关系。3G 为云计算带来数以亿计的宽带移动用户。移动终端的计算能力和存储空间有限，却有很强的联网能力，如果有云计算平台的支撑，移动用户将获得前所未有的服务体验；物联网使用数量惊人的传感器、RFID 和视频监控单元等，采集到极其海量的数据，通过 3G 和宽带互联网进行传输，如果汇聚到云计算设施进行存储和处理，则可以更加迅速、准确、智能、低成本地对物理世界进行管理和控制，大幅提高社会生产力水平和生活质量。

那么，云计算能不能像供水、供电一样，一个国家甚至全世界可以共用一个网络，大家直接按使用收费呢？很明显，软件业在整体动作模式上还远远落后于传统产业，软件业的生产、分发渠道没有水电那么统一和规范，同时最终用户的修改化需求也颇多，需要定制化解决，很难像用水用电那样“千人一面”，所以云计算的问世，就注定了和音乐等各环节自成体系的产业更为相似，需要生产、渠道和消费定制分离。云计算需要基础设施提供商和制造商来做服务器和数据中心，需要掌握下游基础设施资源和上游生态环境的平台提供商来搭建通道，需要迎合最终用户需求的开发商来开发应用服务。

云计算的影响将是深远的，它将彻底改变 IT 产业的架构和运行方式。由此可以预见，高性能计算机、高端服务器、高端存储器和高端处理器的市场将被数量众多、低成本、低能耗和高性价比的云计算硬件市场所占据；绝大多数软件将以服务方式呈现，甚至连大多数游戏都将在“云”里运行；呼叫中心、网络会议中心、智能监控中心、数据交换中心、视频监控中心和销售管理中心等，将越来越向某些云计算设施集中而获取更高的性价比。

放眼远眺，云计算将与网络计算融为一体，实现云计算平台间的互操作和资源共享，实现紧耦合高性能科学计算与松耦合高吞吐量商业计算的融合，使互联网上的主要计算设施融为一个有机整体，即称之为云格。

云计算如此重要，与云计算相关的书籍也应运而生。但由于云计算技术起源于企业界而非学术界，各种技术文献很难寻获齐全，目前还未见到对云计算技术进行全面、深入剖析的教科书。本书便是在此情况下对云计算的入门及实战进行深入剖析，具有如下特点：

- (1) 涉及内容广泛、全面，深入浅出。
- (2) 由入门到实践概括地介绍云计算，做到理论与实践相结合。
- (3) 内容新颖，图文并茂。

本书共 7 章，分别介绍如下。

第 1 章：介绍了云计算绪论，主要包括云计算的定义、产生背景、发展史、层次结构等内容。

第 2 章：介绍了 Google 云计算实现，主要包括 Google 整体架构说明、Google 文件系统 GFS、Google 计算架构及 GFS 应用实例等内容。

第 3 章：介绍了 Hadoop 分布式文件系统，主要包括 Hadoop 分布式文件系统 HDFS、Hadoop 分布式数据处理 MapReduce 及 Hadoop 分布式存储数据 HBase 等内容。

第 4 章：介绍了 Amazon 云计算，主要包括 Dynamo 存储架构、EC2 云计算、存储服务 S3 及数据库服务 RDS 等内容。

第 5 章：介绍了 VMware 云计算，主要包括 VMware 概述、ESX/ESXi、vCenter 及 DRS Cluster 等内容。

第 6 章：介绍了 Windows Azure 云计算，主要包括 Windows Azure 概述、Windows Azure 平台的体系结构、Windows Azure 操作系统及 SQL Azure 体验等内容。

第 7 章：介绍了云计算其他应用软件，主要包括 Hive、Pig、Cassandra 及 Zookeeper 等。

本书主要由张德丰主编，此外参加编写的还有丁伟雄、雷晓平、李娅、杨文茵、何正风、赵新芬、赵书梅、栾颖、刘志为、周灵、余智豪、赵书兰和崔如春。

由于作者的水平有限，加之时间紧迫，书中难免会存在不足之处，敬请广大读者批评指正。

编者

# 目 录

第 1 章 云计算绪论 .....	1
1.1 云计算概述 .....	1
1.1.1 云计算的定义 .....	1
1.1.2 云计算产生的背景.....	2
1.1.3 云时代谁是主角 .....	2
1.1.4 云计算的特征 .....	4
1.1.5 云计算的发展史 .....	5
1.1.6 云计算的服务层次.....	6
1.1.7 云计算的服务形式.....	7
1.1.8 云计算的实现机制.....	9
1.1.9 云计算延伸 .....	10
1.1.10 云计算研究方向 .....	12
1.1.11 云计算发展趋势 .....	14
1.2 云计算的发展环境 .....	15
1.2.1 云计算与对等网的关系.....	15
1.2.2 云计算与网格计算关系.....	18
1.2.3 云计算与物联网关系.....	21
1.2.4 云计算与 3G 关系.....	23
1.3 云计算研究热点 .....	24
1.3.1 研究云计算体系结构.....	25
1.3.2 研究云计算关键技术.....	27
1.3.3 研究云计算支撑平台.....	29
1.3.4 研究云计算的安全性.....	34
1.4 云标准 .....	37
1.5 云计算的生命周期 .....	39
1.6 我国云计算中心的问题与注意事项.....	41
第 2 章 Google 云计算实现 .....	42
2.1 Google 整体架构概述.....	42
2.1.1 Google 技术整体构架.....	42
2.1.2 Google 外部与内部技术层次介绍.....	43
2.2 Google 文件系统 GFS .....	44
2.2.1 文件系统架构 .....	45
2.2.2 GFS 的特点 .....	46

---

2.2.3	文件系统的容错性.....	47
2.2.4	系统管理技术 .....	47
2.3	Google 计算架构.....	48
2.3.1	Google 并行计算架构 MapReduce.....	48
2.3.2	Google 分布式锁服务 Chubby .....	50
2.3.3	Google 分布式结构化数据库 BigTable .....	50
2.4	GFS 应用实例 .....	53
2.5	Google 应用引擎.....	54
2.5.1	Google 应用引擎概述 .....	55
2.5.2	Google 应用引擎环境 .....	56
2.5.3	Google 应用引擎服务 .....	57
2.5.4	Google 应用引擎实例 .....	58
2.6	Hadoop 云计算概述.....	68
2.6.1	Hadoop 的构架介绍 .....	69
2.6.2	Hadoop 群族.....	71
2.6.3	Hadoop 在 Windows 中安装.....	72
2.6.4	Hadoop 的应用 .....	80
2.7	Hadoop 云计算系统操作 .....	81
2.7.1	多节点 Hadoop 云计算系统设置 .....	81
2.7.2	多节点 Hadoop 设置实例 .....	83
2.7.3	启动 Hadoop 云计算系统 .....	86
2.8	Hadoop 的 Avatar 机制 .....	88
2.8.1	Avatar 系统架构.....	88
2.8.2	Avatar 元数据同步机制.....	89
2.8.3	切换故障过程 .....	91
2.8.4	Avatar 运行流程.....	92
2.8.5	切换 Avatar 故障流程 .....	95
2.9	Hadoop 应用开发.....	99
第 3 章	Hadoop 分布式文件系统 .....	109
3.1	Hadoop 分布式文件系统 HDFS.....	109
3.1.1	设计前提与目标 .....	109
3.1.2	HDFS 体系结构 .....	110
3.1.3	HDFS 的可靠性措施 .....	111
3.1.4	访问接口 .....	115
3.1.5	HDFS 常用命令 .....	116
3.1.6	HDFS 应用实例 .....	119
3.2	Hadoop 分布式数据处理 MapReduce.....	126
3.2.1	MapReduce 云计算模型 .....	126
3.2.2	MapReduce 架构结构 .....	127

3.2.3	MapReduce 云计算应用实例 .....	129
3.2.4	并行随机发生器算法的 MapReduce 实现.....	141
3.3	MapReduce 的进一步功能 .....	143
3.3.1	MapReduce 的进一步功能概述 .....	143
3.3.2	定制的数据类型 .....	144
3.3.3	定制的输出/输入形式 .....	146
3.3.4	连接多数据源 .....	151
3.3.5	传递与使用全局参数/数据文件.....	157
3.3.6	连接与访问关系数据库.....	160
3.4	Hadoop 分布式存储数据 HBase .....	162
3.4.1	HBase 的系统框架 .....	162
3.4.2	HBase 的模型 .....	164
3.4.3	HBase 的安裝配置 .....	168
3.4.4	HBase 的优缺点 .....	175
3.4.5	HBase 应用实例 .....	175
第 4 章	Amazon 云计算 .....	180
4.1	Dynamo 存储架构.....	180
4.1.1	Dynamo 的地位.....	180
4.1.2	Dynamo 数据技术.....	181
4.1.3	Dynamo 主要技术.....	182
4.2	EC2 云计算 .....	189
4.2.1	EC2 的主要特性 .....	189
4.2.2	EC2 的基本架构 .....	189
4.2.3	EC2 的相关技术 .....	190
4.2.4	EC2 安全与容错性 .....	192
4.3	存储服务 S3 .....	193
4.3.1	S3 基本概念 .....	193
4.3.2	S3 一致性 .....	195
4.3.3	S3 安全措施 .....	196
4.3.4	简单队列服务 .....	198
4.3.5	Simple DB .....	199
4.3.6	S3 支持的程序语言种类 .....	202
4.4	数据库服务 RDS.....	204
4.4.1	SQL 与 NoSQL 数据库比较.....	204
4.4.2	RDS 数据库原理及使用 .....	205
4.5	AWS 应用实例.....	206
4.6	Eucalyptus.....	208
4.6.1	Eucalyptus 概述.....	208
4.6.2	Eucalyptus 体系结构.....	209

---

4.6.3	Eucalyptus 高级特点.....	210
4.6.4	Eucalyptus 的安装.....	211
4.6.5	Eucalyptus 应用实例.....	213
<b>第 5 章</b>	<b>VMware 云计算 .....</b>	<b>216</b>
5.1	VMware 概述 .....	216
5.1.1	VMware 主要特点 .....	216
5.1.2	VMware 产品类型 .....	217
5.1.3	VMware 主要功能 .....	217
5.1.4	VMware 的 3 种工作模式 .....	217
5.1.5	安装 VMware Player .....	218
5.2	ESX/ESXi .....	220
5.2.1	ESX/ESXi 定义 .....	220
5.2.2	安装 ESX 与配置 .....	224
5.2.3	ESX 应用实例 .....	231
5.3	vCenter .....	232
5.3.1	vCenter 概述 .....	232
5.3.2	vCenter 安全性与可靠性措施 .....	235
5.3.3	安装 vCenter .....	237
5.4	vCloud Service Director .....	243
5.4.1	vCloud Service Director 功能 .....	243
5.4.2	vCloud Service Director 特点 .....	245
5.4.3	VMware 的网络和存储虚拟化 .....	245
5.5	DRS Cluster .....	248
5.5.1	DRS Cluster 定义 .....	248
5.5.2	将 HA DRS Cluster 和主机添加到 vCenter .....	248
5.6	VMware HA .....	251
5.6.1	VMware HA 概述 .....	251
5.6.2	VMware HA 的完整过程 .....	252
<b>第 6 章</b>	<b>Windows Azure 云计算 .....</b>	<b>258</b>
6.1	Windows Azure 概述 .....	258
6.1.1	Windows Azure .....	258
6.1.2	SQL Azure .....	259
6.1.3	Windows Azure AppFabric .....	260
6.1.4	Windows Azure Connect .....	262
6.2	Windows Azure 平台的体系结构 .....	262
6.2.1	Windows Azure 创建项目 .....	262
6.2.2	Azure 应用和服务 .....	266
6.3	Windows Azure 操作系统 .....	267

---

6.3.1	Windows Azure OS 指南 .....	268
6.3.2	Windows Azure 生命周期.....	269
6.3.3	Windows Azure 安全性 .....	271
6.3.4	Windows Azure 的可用性.....	273
6.3.5	Windows Azure 的完整性.....	273
6.3.6	Windows Azure 的可靠性.....	274
6.3.7	Windows Azure 数据隔离.....	274
6.3.8	Windows Azure 存储服务.....	276
6.4	Windows Azure 服务在企业中的应用.....	278
6.5	SQL Azure 体验 .....	288
6.5.1	如何申请创建一个 SQL Azure 账户.....	289
6.5.2	如何创建一个云端 SQL Azure 服务器实例.....	291
6.5.3	如何设置防火墙规则.....	293
6.5.4	如何创建一个云端 SQL Azure 数据库.....	293
6.5.5	如何创建一个云端 SQL Azure 服务器登录账户.....	296
6.5.6	如何修改登录账户密码.....	296
6.5.7	如何进行本地管理.....	296
6.6	Windows Azure 应用实例 .....	298
6.6.1	Windows Azure Table 架构 .....	298
6.6.2	Windows Azure Table 数据模型.....	299
6.6.3	实现图片管理 .....	299
<b>第 7 章</b>	<b>云计算其他应用软件 .....</b>	<b>306</b>
7.1	SaaS 软件服务应用 .....	306
7.2	Hive.....	307
7.2.1	Hive 组成.....	307
7.2.2	Hive 安装.....	308
7.2.3	Hive 结构解析.....	311
7.2.4	HiveQL 的使用 .....	314
7.2.5	Hive 应用实例.....	321
7.2.6	Hive 实战.....	324
7.3	Pig.....	329
7.3.1	Pig 基本框架 .....	329
7.3.2	Pig 安装.....	330
7.3.3	Pig 的使用 .....	330
7.3.4	Pig 编程语言 .....	334
7.3.5	Pig 应用实例 .....	342
7.4	Cassandra .....	346
7.4.1	Cassandra 概述.....	346
7.4.2	Cassandra 操作过程 .....	350

---

7.4.3	Cassandra 安装 .....	354
7.4.4	Cassandra 使用 .....	356
7.4.5	Cassandra 应用实例 .....	361
7.4.6	Cassandra 与 MapReduce 相结合 .....	363
7.5	Zookeeper .....	366
7.5.1	Zookeeper 概述 .....	366
7.5.2	Zookeeper 工作原理 .....	368
7.5.3	Zookeeper 安装 .....	372
7.5.4	Zookeeper API 接口 .....	374
7.5.5	Zookeeper 编程 .....	374
7.5.6	Zookeeper 应用实例 .....	376
附录 A	参考资料 .....	385

# 第 1 章 云计算绪论

很少有一种技术能够像云计算这样，在短短的两三年间就产生巨大的影响力。Google、Amazon、IBM 和微软等 IT 巨头们以前所未有的速度和规模推动云计算技术和产品的普及，相关的一些学术活动广泛展开。

## 1.1 云计算概述

云计算即为新的 Web 2.0，一种既有技术上的市场绽放。就像以前人们在自己的网站上放一点 Ajax 就宣称自己为 Web 2.0 一样，云计算是一个新的概念。积极的一面是，Web 2.0 最终抓住了主流眼球，同样，云计算概念也会改变人们的思维，最终派生出各种各样的概念，如托管服务、ASP、网格计算、软件作为服务、平台作为服务、任何东西作为服务。从消费者的角度看，SaaS 是云计算的一种，然而行业内的人必须明白其含义。简单地说，云计算就是 SaaS 的升华。

### 1.1.1 云计算的定义

云计算（Cloud Computing）是在 2007 年第三季度才诞生的新名词，但仅过了半年多，其受到关注的程度就超过了网格计算（Grid Computing），如图 1-1 所示。云计算是一种基于互联网的计算方式，通过这种方式，共享的软硬件资源和信息可以按需要提供给计算机和其他设备。“云”其实是网络、互联网的另一种说法。云计算的核心思想是将大量用网络连接的计算资源统一管理和调度，构成一个计算资源池，向用户按需服务。提供资源的网络被称为“云”。狭义的云计算即指 IT 基础设施的交付和使用模式，指通过网络以按需、易扩展的方式获取所需资源；广义的云计算指服务的交付和使用模式，指通过网络以按需、易扩展的方式获取所需服务。这种服务可以是 IT 和软件、互联网相关，也可是其他服务。

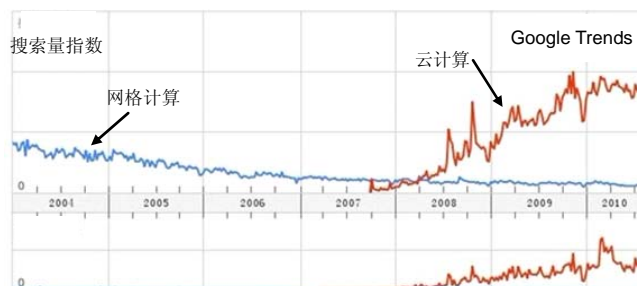


图 1-1 云计算和网格计算在 Google 中的搜索趋势图

### 1.1.2 云计算产生的背景

有人说云计算是技术革命的产物，有人说云计算只不过是已有技术的重新包装，是设备厂商或软件厂商“换汤不换药”的一种商业策略。而笔者认为，云计算的发展是需求推动、技术进步及商业模式转换共同作用下的结果。

(1) 需求是云计算的动力。

IT 设施要成为社会基础设施，现在面临高成本的瓶颈，这些成本至少包括人力成本、资金成本、时间成本、应用成本、环境成本。云计算带来的益处是显而易见的：用户不需要专门的 IT 团队，不需要购买、维护、安放有形的 IT 产品，可以低成本、高效率、随时、快捷地按需使用 IT 服务；云计算服务提供商可以极大提高资源（硬件、软件、空间、人力、物力、资源等）的利用率和业务响应速度，有效聚合产业链。

(2) 技术是云计算发展的基础。

云计算自身核心技术的发展，如硬件技术、虚拟化技术（计算虚拟化、网络虚拟化、存储虚拟化、桌面虚拟化、应用虚拟化）、海量存储技术、分布式并行计算、多用户构架、自动管理与部署；云计算赖以存在的移动互联网技术的发展，如高速、大容量的网络，无处不在的接入，灵活多样的终端，集约化的数据中心 Web 技术。

(3) 商业模式是云计算的内在要求，是用户需求的外在体现，并且云计算技术为这种特定商业模式提供了现实可能性。

从商业模式的角度看，云计算的主要特征是以网络为中心、以服务为产品形态、按需使用与付费，这些特征分别对应于传统的用户自建基础设施、购买有形产品或介质（含 Licence）、一次性买断。

纯粹从技术角度看，云计算是很多技术自然发展、精心优化与组合的产物，是这些技术的集大成者；另一方面，如果同时考虑到商业模式，那么可断言，云计算将给整个社会的信息化带来革命性的改变。所以，在此绝不能离开技术谈云计算，否则有“忽悠”之嫌；也不能离开商业模式谈云计算，否则云计算就是“无源之水，无本之木”。

### 1.1.3 云时代谁是主角

纵观整个 ICT 产业发展历程，每一次计算模式的变革都会引发一场产业变革，同时也会造就一批“明星”厂商。主机时代，IBM 风光无限，称霸一时；互联网时代，微软、Intel 即为名副其实的行业主导者。然而，云计算时代，谁才是主角？

#### 1. Google

Google 是最早提倡和实践云计算技术的企业之一，其互联网搜索服务建立在云计算基础架构之上。经过多年的发展，Google 云计算技术逐渐成熟，针对自身特点建立了一套极其有效的商业模式与产品、服务组合。2011 年 8 月，Google 以 6820 万美元收购企业级 IP 通信解决方案提供商 GIPS，并于同年召开了“Google I/O 开发者大会”，发布了以企业级

Google App 为核心的云计算产品。种种事实表明，Google 云计算目标并不只在于个人用户，它的野心在于覆盖从个人用户至企业用户的广大空间。Google 还积极与其他云计算企业合作。作为全球最具有影响力的高科技企业之一的 Google，正以一种先行者的姿态拥抱云计算时代的到来。

Google 标志如图 1-2 所示。

## 2. IBM

IBM 是云计算领域中名副其实的巨头，2007 年高调启动“蓝云”计划，推出一系列云计算产品。2008 年，IBM 在云计算领域的累计投入超过了 10 亿美元，将其云计算产品和服务扩展到亚洲、欧洲、非洲、美洲市场。为了进一步抢占全球云计算市场，从 2009 年开始 IBM 加大了在云计算领域上的投入。有消息透露，IBM 将投资 200 亿美元进行并购、开发云计算终端、推出网络软件……摆出了一副势在必得的架势。IBM 在 IaaS、PaaS、SaaS 3 个层面都有方案推出，公有云、私有云、混合云一应俱全。近两年来，IBM 的“智慧”战略如火如荼，智慧地球、智慧城市、智慧通信、智慧医疗……一切都是智慧的。IBM 智慧的云计算也是其智慧战略中的重要组成部分，其云智慧正不断向云计算领域延伸。

IBM 标志如图 1-3 所示。



图 1-2 Google 标志



图 1-3 IBM 标志

## 3. 微软

云+端、软件+服务是对微软云计算的最佳诠释，其云计算平台 Windows Azure 被认为是 Windows NT 之后，16 年来最重要的产品，关乎微软的未来。微软 CEO 史蒂夫·鲍尔默多次公开表示，云计算是微软的又一次机遇。与其他云计算厂商相比，微软在用户上有着明显的优势。微软的操作系统和操作习惯，不管是在个人用户还是企业用户中都有很广泛的影响力。微软可以借助这些优势迅速推广其云计算产品和服务，微软基于云计算的解决方案正在得到越来越广泛的应用。近日有消息称，微软云计算用户已经超过 Google App 用户。这表明，微软云计算产品的势力范围正在逐步扩大，将对产业链上下游产生深远影响。

微软标志如图 1-4 所示。



图 1-4 微软标志

### 1.1.4 云计算的特征

之所以称为“云”，是因为它在某些方面具有现实中云的特征。

- ❑ 云一般都较大。
- ❑ 云的规模可以动态伸缩，它的边界是模糊的。
- ❑ 云在空中飘忽不定，无法也无须确定它的具体位置，但它确实存在于某处。同时还因为云计算的鼻祖之一亚马逊公司将大家曾经称为网格计算的东西，取了一个新名为“弹性计算云”（Elastic Computing Cloud），并取得了商业上的成功。

有人将这种模式比作从单台发电机供电模式转向电厂集中供电的模式，这意味着计算能力也可以作为一种商品进行流通，就像天然气、水和电一样，使用方便，费用低廉。最大的不同在于，它是通过互联网进行传递的。

云计算是并行计算（Parallel Computing）、分布式计算（Distributed Computing）及网格计算（Grid Computing）的发展，或者说是这些计算科学概念的商业实现。云计算是虚拟化（Virtualization）、效用计算（Utility Computing）、将基础设施作为服务 IaaS（Infrastructure as a Service）、将平台作为服务 PaaS（Platform as a Service）和将软件作为服务 SaaS（Software as a Service）等概念混合演进并跃升的结果。从研究现状上看，云计算具有以下特点：

（1）超大规模。“云”具有相当的规模，Google 云计算已经拥有 100 多万台服务器，亚马逊、IBM、微软和 YAHOO 等公司的“云”均拥有几十万台服务器等。“云”能赋予用户前所未有的计算能力。

（2）虚拟化。云计算支持用户随时、随地使用各种终端获取服务。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，但实际上用户无须了解应用运行的具体位置，只需要一台笔记本电脑或一个 PDA，就可以通过网络服务来获取各种能力超强的服务。

（3）提高设备计算能力。云计算把大量计算资源集中到一个公共资源池中，通过多主租用的方式共享计算资源。虽然单个用户在云计算平台获得的服务水平受到网络带宽等各因素影响，未必获得优于本地主机所提供的服务，但是从整个社会资源的角度而言，整体的资源调控降低了部分地区峰值荷载，提高了部分荒废的主机的运行率，从而提高了资源的利用率。

（4）高可靠性。“云”使用了数据多容错性、计算节点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加可靠。

(5) 减少设备依赖性。虚拟化层将云平台上方的应用软件和下方的基础设备隔离开来。技术设备的维护者无法看到设备中运行的具体应用。同时对软件层的用户而言基础设备层是透明的，用户只能看到虚拟化层中虚拟出来的各类设备。这种架构减少了设备依赖性，也为动态的资源配置提供可能。

(6) 通用性。云计算不针对特定的应用，在“云”的技术支撑下可以构造出千变万化的应用，同一片“云”可以同时支撑不同的运行程序。

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

(8) 弹性服务。云平台管理软件将整合的计算资源根据应用访问的具体情况进行动态调整，包括增大或减少资源的要求。因此云计算对于非恒定需求，如对需求波动很大、阶段性需求等，具有非常好的应用效果。在云计算环境中，既可以对规律性需求通过事先预测事先分配，也可根据事先设定的规则进行实时公告调整。弹性的云服务可帮助用户在任意时间得到满足需求的计算资源。

(9) 按需服务。“云”是一个庞大的资源池，用户按需购买，就像自来水、电和天然气那样计费。

(10) 极其廉价。“云”的特殊容错措施使得可以采用极其低价的节点来构成云；“云”的自动化管理使数据中心管理成本大大降低；“云”的公用性和通用性使资源的利用率大幅提升；“云”设施可以构建在电力资源丰富的地区，从而大大降低能源成本，因此“云”具有前所未有的性价比。

Google 中国区前总裁李开复声称：Google 每年投入约 16 亿美元构建云计算数据中心，所获得的能力相当于使用传统技术投入 640 亿美元，节省了约 40 倍的成本。因此，用户可以充分享受“云”的低成本优势，需要时，花费几百美元、一天时间就能完成以前需要数万美元、数月时间才能完成的数据处理任务。

### 1.1.5 云计算的发展史

云计算的起源要先从互联网演进讲起，如图 1-5 所示为云计算的演进与由来。云计算从根本上改变了原有的互联网结构，将计算能力从个人终端向服务端靠拢，弱化了端的概念，提高了计算资源的整体利用率。在量化计算资源的基础上，云计算实现了商业模式由设置向服务进化的过程。更令人满意的是，随着全球互联网的发展，云计算被赋予了更为广泛的定义：从连接计算资源到连接所有的人和机器设置，计算能力也将进一步智能化。



图 1-5 云计算的演进与由来

云计算的发展过程如下：

1983 年，太阳电脑（Sun Microsystems）提出“网络是电脑（The Network is the Computer）”；2006 年 3 月，亚马逊（Amazon）推出弹性计算云（Elastic Compute Cloud, EC2）服务。

2006 年 8 月 9 日，Google 首席执行官埃里克·施密特（Eric Schmidt）在搜索引擎大会（SES San Jose 2006）首次提出“云计算”（Cloud Computing）的概念。Google“云端计算”源于 Google 工程师克里斯托弗·比希利亚的 Google 101 项目。

2007 年 10 月，Google 与 IBM 开始在美国大学校园推广云计算的计划，包括卡内基·梅隆大学、麻省理工学院、斯坦福大学、加州大学柏克莱分校及马里兰大学等，这项计划希望能降低分布式计算技术在学术研究方面的成本，并为这些大学提供相关的软硬件设备及技术支持（包括数百台个人电脑及 BladeCenter 与 System X 服务器，这些计算平台将提供 1600 个处理器，支持包括 Linux、Xen、Hadoop 等开放源代码平台）。而学生则可以通过网络开发各项以大规模计算为基础的研究计划。

2008 年 1 月 30 日，Google 宣布在台湾启动“云计算学术计划”，将与台湾大学、台湾交通大学等学校合作，将这种先进的大规模、快速计算技术推广到校园。

2008 年 2 月 1 日，IBM（NYSE: IBM）宣布将在中国无锡太湖新城科教产业园为中国的软件公司建立全球第一个云计算中心（Cloud Computing Center）。

2008 年 7 月 29 日，雅虎、惠普和英特尔宣布一项涵盖美国、德国和新加坡的联合研究计划，推出云计算研究测试床，进而推进云计算。该计划要与合作伙伴创建 6 个数据中心作为研究试验平台，每个数据中心配置 1400~4000 个处理器。这些合作伙伴包括新加坡资讯通信发展管理局、德国卡尔斯鲁厄大学 Steinbuch 计算中心、美国伊利诺伊大学香槟分校、英特尔研究院、惠普实验室和雅虎。

2008 年 8 月 3 日，美国专利商标局网站信息显示，戴尔正在申请“云计算（Cloud Computing）”商标，此举旨在加强对这一未来可能重塑技术。

2010 年 3 月 5 日，Novell 与云安全联盟（CSA）共同宣布一项供应商中立计划，名为“可信任云计算计划（Trusted Cloud Initiative）”。

2010 年 7 月，美国国家航空航天局和包括 Rackspace、AMD、Intel、戴尔等支持厂商共同宣布 OpenStack 开放源代码计划，微软在 2010 年 10 月表示支持 OpenStack 与 Windows Server 2008 R2 的集成；而 Ubuntu 已把 OpenStack 加至 11.04 版本中。

2011 年 2 月，思科系统正式加入 OpenStack，重点研制 OpenStack 的网络服务。

2011 年 10 月 20 日，“盛大云”宣布旗下产品 MongoIC 正式对外开放，这是中国第一家专业的 MongoDB 云服务，也是全球第一家支持数据库恢复的 MongoDB 云服务。

### 1.1.6 云计算的服务层次

在云计算中，根据其服务集合所提供的服务类型，全部云计算服务集合被划分成应用层、平台层、基础设施层和虚拟化层 4 个层次，每一层都对应着一个子服务集合。图 1-6 所示为云计算的服务层次效果。