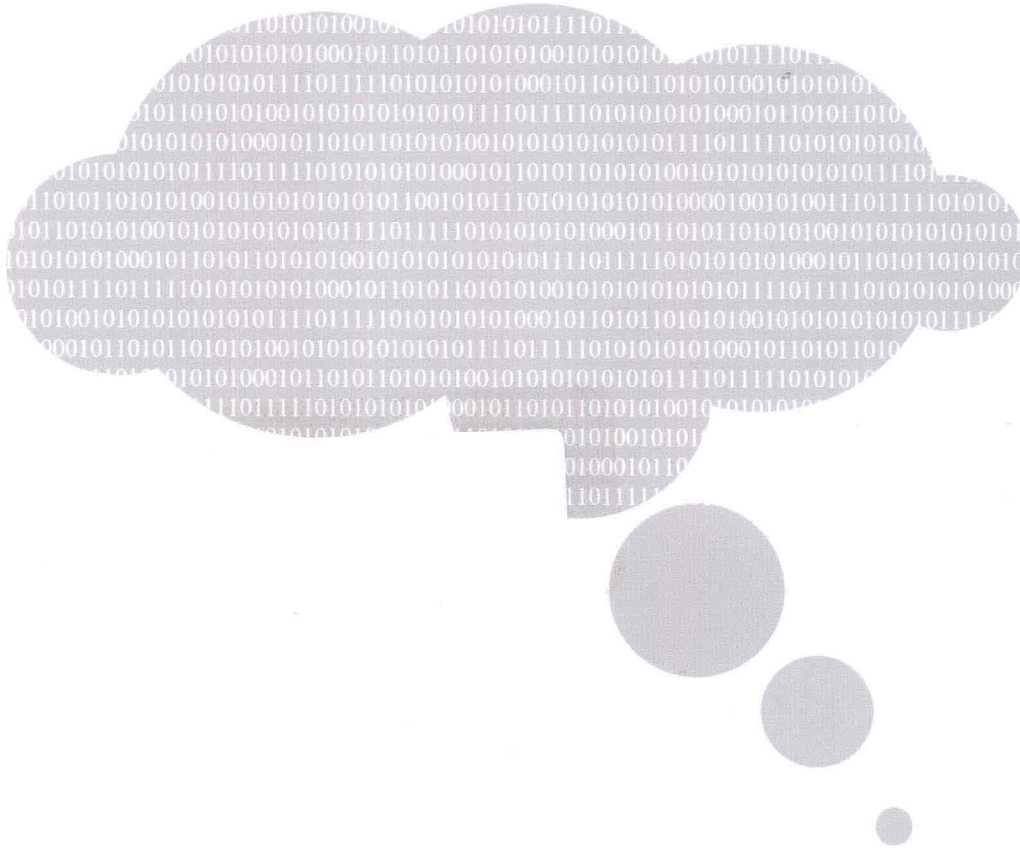
云时代的大数据

周　品　编著



電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

云时代的大数据

周　品　编著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

本书以云计算为基石，从概念、研究、应用角度出发，系统地介绍了数据爆炸时代下的大数据。首先介绍了云计算及大数据的基础知识，让读者对云计算及大数据有概要认识；然后根据需要介绍了 Hadoop 软件下的 MapReduce、HDFS 及 HBase 这几个组件；接着全面、系统地介绍了云时代下的大数据，主要包括大数据的链接、聚类、项集、系统、相似项挖掘及数据量化等内容，让读者挖掘云时代大数据体系下的效益、价值及研究方向。

本书适合于学习、研究、应用大数据的读者阅读。

图书在版编目（CIP）数据

云时代的大数据 / 周品编著. —北京：电子工业出版社，2013.10
ISBN 978-7-121-21644-2

Ⅰ. ①云… Ⅱ. ①周… Ⅲ. ①计算机网络—研究 Ⅳ. ①TP393

中国版本图书馆 CIP 数据核字（2013）第 240129 号

策划编辑：陈韦凯
责任编辑：毕军志
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
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：21.75　字数：556.8 千字
印　　次：2013 年 10 月第 1 次印刷
印　　数：3 500 册　　定价：58.00 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。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前　言

几年之内，云计算已从新兴技术发展成为当今的热点技术。从 2003 年谷歌公开发布的核心文件到 2006 年 Amazon EC2（亚马逊弹性计算云）的商业化应用，再到美国电信巨头 AT&T（美国电话电报公司）推出的 Synaptic Hosting（动态托管）服务，云计算从节约成本的工具到盈利的推动器，从 ISP（网络服务提供商）到电信企业，已经成功地从内置的 IT 系统演变成公共的服务。

随着云时代的来临，大数据（Big data）也吸引了越来越多的关注。《著云台》[①] 的分析师团队认为，大数据通常用来形容一个公司创造的大量非结构化和半结构化数据，这些数据在下载到关系数据库中用于分析时会花费过多的时间和金钱。大数据分析常和云计算联系到一起，因为实时的大型数据集分析需要像 MapReduce 一样的框架来向数十、数百甚至数千台计算机分配工作。

“大数据”这个术语最早期的引用可追溯到 Apache 的开源项目 Nutch。当时，大数据用来描述为更新网络搜索索引需要同时进行批量处理或分析的大量数据集。随着谷歌 MapReduce 和 Google File System（GFS）的发布，大数据不再仅用来描述大量的数据，还涵盖了处理数据的速度。

在当今的 IT 行业中都需要对数据进行分析，而数据分析都需要数据源，大数据尤甚。互联网公司通过搜索引擎、访问记录、App 追踪等技术手段可以获得大量的用户浏览信息，但这些信息的收集、存储、提取、访问等环节都不可能向大众公开，相关数据的使用规则目前还缺乏法律规范。对普通人而言，获得公开、免费、准确的数据来源似乎成为一种奢望，但企业和政府的数据公开的步伐已经迈出。各行各业都需要大数据，如医疗上的各种疾病数据，农业上的作物、天气、病虫害、土壤资料等数据，工业制造上的原材料、加工流程、设备信息、产品规格等数据，金融行业的客户资料、金融产品等数据，教育领域的学生、学校、教师、教材等数据，国防领域的卫星、海域等数据，环境保护中的空气污染物、水源质量分析等实时数据……不论政府、企业还是个人，对大数据的需求也涉及经济社会的各个方面。

互联网和移动互联网已经给电信、零售、媒体等行业带来了深刻变革，如果进入大数据时代，那么还有更多行业会迎接洗礼。目前智能制造、互联网金融、数字化诊疗已经崭露头角。个人用户对大数据的需求可能带来数据接收方式的变化，各类智能终端将再次迎来发展机遇。除了功能越来越强大的智能手机之外，眼镜、汽车、手表，甚至自行车都有可能成为接收数据的新型智能终端。

根据云时代的大数据发展趋势，笔者编著了本教材，让读者认识到什么是云，什么是大数据，以及云与大数据的关系，在各企业领域中怎样应用云时代的大数据。本书主要内容有：

第 1 章：介绍了云时代概述，主要包括“云”基本介绍、云产生的背景、云计算特点及云计算的关键性技术等内容。

第 2 章：介绍了大数据概述，主要包括大数据基本概念、大数据的发展趋势、大数据的挑战、现状与展望及大数据行业应用和未来热点等内容。

---

① 《中国云》平台与受众营销联盟云生态系统，云时代云计算概念领军品牌商标之一。

第 3 章：介绍了数据挖掘，主要包括数据挖掘的定义、起源、功能、实现方法、应用及哈希函数等内容。

第 4 章：介绍了数据量化，主要包括量化分析元素、量化质量分析规划及高级量化分析等内容。

第 5 章：介绍了大规模文件系统 MapReduce，主要包括分布式文件系统、MapReduce 模型、MapReduce 使用算法及 MapReduce 实现机制等内容。

第 6 章：介绍了相似项挖掘，主要包括近邻搜索的应用、最小哈希及距离测试等内容。

第 7 章：介绍了 HDFS 存储海量数据，主要包括 HDFS 简介、HDFS 存取机制及 HDFS 管理操作等内容。

第 8 章：介绍了 HBase 存储百科数据，主要包括 HBase 基本特征、系统框架、HBase 的基本接口及 HBase 数据模型等内容。

第 9 章：介绍了大数据链接分析，主要包括链接分析中的数据采集研究、PageRank 及搜索引擎研究等内容。

第 10 章：介绍了聚类，主要包括聚类概述、聚类技术、层次聚类用 CURE 算法等内容。

第 11 章：介绍了项集与系统，主要包括项集规则、单调性、二元组计数及推荐模型系统等内容。

本书适用于云计算及大数据初、中、高级读者使用，也可作为研究大数据相关专业研究人员的参考资料。

由于时间仓促，加之作者水平有限，所以错误和疏漏之处在所难免。在此，诚恳地期望得到各领域的专家和广大读者的批评指正。

编著者

2013 年 6 月

目　录

第 1 章　云时代概述

什么是云时代？云时代是指云计算时代，云计算（Cloud Computing）是分布式处理（Distributed Computing）、并行处理（Parallel Computing）和网格计算（Grid Computing）的发展，或者说是这些计算机科学概念的商业实现，这将是一个时代的来临。

1.1 “云”概述

“云”即是计算机群，每一群包括了几十万台，甚至上百万台计算机。“云”的好处在于，计算机可以随时更新，保证“云”长生不老。谷歌就有好几个这样的“云”，如微软、雅虎、亚马逊（Amazon）也有或正在建设这样的“云”。届时，只需要一台能上网的计算机，无须关心存储或计算发生在哪朵“云”上，一旦有需要，可以在任何地点用任何设备，如计算机、手机等，快速地计算和找到所需的资料，再也不用担心资料丢失。

这是一种革命性的举措，打个比方，这就好比是从古老的单台发电机模式转向了电厂集中供电的模式。其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、水电一样，取用方便，费用低廉。最大的不同在于，它是通过互联网进行传输的。云计算的蓝图已经呼之欲出：在未来，只需要一台笔记本电脑或者一部手机，就可以通过网络服务来实现我们需要的一切，甚至包括超级计算这样的任务。从这个角度而言，最终用户才是云计算的真正拥有者。云计算的应用包含这样的一种思想，把力量联合起来，给其中的每一个成员使用。从最根本的意义来说，云计算就是利用互联网上的软件和数据的能力。

图 1-1 所示为云时代效果图。

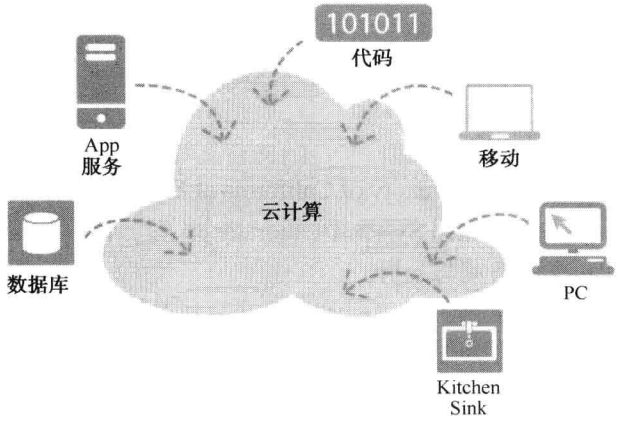


图 1-1　云时代效果图

1.1.1 什么是云计算

云计算是多种技术混合演进的结果，包括网格计算、效用计算、虚拟化技术、Web Services、SOA 等，上述热点技术的融合发展将为我国的 IT 产业，特别是软件服务业带来影响广泛的变革。

自 2007 年以来云计算逐渐成为业界的认可和推崇的技术热点。众多国内外厂商围绕云计算开发出大量的产品，同时，越来越多的互联网应用开始尝试使用云服务构建基于云计算的解决方案，以及各大企业的关注热点。官方的国际标准化组织以及多个国际协会组织近两年来纷纷启动了云计算相关标准化工作，我国相关标准化组织也启动了云计算的标准研究及制定。

云计算目前在不同的组织、机构、企业都有定义，不同组织的定义往往关注于技术的特定方面。

1. 维基百科

云计算将 IT 相关的能力以服务的方式提供给用户，允许用户在不了解提供服务的技术、没有相关知识以及设备操作能力的情况下，通过 Interent 获取需要的服务。

2. 国际标准化组织 ISO/IEC JTC1 的云计算相关报告（2009 年）

（1）云计算是提取的、高级的可升级的池，是能够为终端用户提供主机应用和通过消费买单的管理计算基础设施。

（2）动态可升级的计算风格和通常虚拟化的资源在因特网上作为一个服务提供。用户不需要精通或者控制支撑他们的“云”中的技术基础设施。

（3）云计算是新出现的共享基础设施，将大型池系统连接在一起以提供 IT 服务。

（4）暂时存储在因特网服务器上的信息的范围和关于客户的暂时高速缓冲存储器，包括台式计算机、娱乐中心、笔记本、掌上笔记本等（组织可升级的、坚固的、包含隐私的客户机云计算，IEEE（Institute of Electrical and Electronics Engineers，电气和电子工程师协会）因特网计算）。

从业务的角度看，云计算提供了 IT 基础设施和环境以开发/提供主机/运行服务和应用，在需要应用时，即时购买作为一个服务。而且，从用户的角度，云计算提供资源和服务以存储数据和运行应用，在任何设备、任何时间、任何地点，作为一个服务。现在云计算的用法正扩展到众多明确的领域，包括网络服务、移动服务、媒体服务等。

3. 美国加州大学伯克利分校

美国加州大学伯克利分校（University of California at Berkeley）——可靠适应性分布式系统实验室（Reliable Adaptive Distributed Systems Laboratory）对云计算在技术、商业应用中的现状和将来做了比较详细和科学的研究与分析。

他们认为云计算是：在互联网上以“服务”形式交付的应用程序，以及提供和支持这些服务的数据中心（Data Center，包括硬件和软件系统）。

这些服务叫 SaaS（Software as a Service，软件即服务），是数据中心里的软件总称，可称为（一个）云。

当某个云通过支付购买（Pay-as-you-go）的方式被使用时，就叫做公共云（Public Cloud）。这样，服务就被“销售”，这种方式叫做效用计算（Utility Computing）。

相对公共云，我们把公司、组织机构之间内部使用的数据中心叫做私有云（Private Cloud）。通常私有云不对公众开放。

1.1.2 给云计算一个说法

云计算是当前信息行业比较流行的话题，有关它的定义虽不是众说纷纭，但并非完全一致。总体上讲，云计算是一种计算行为或技术风格，特点是在互联网上提供一种动态可扩展的虚拟资源服务。为了满足这种动态可扩展性的要求，云计算服务商必须建立和依靠大型数据中心，它们通常分布在一个国家的各个地区，甚至世界其他国家和地区。可以想象，这样的计算网络有多大、多复杂和多昂贵。云计算中的“云”字是相对互联网而言的，用以比喻互联网的复杂结构。

过去的蒸汽机、信息工业化和互联网的普及给人类社会带来了巨大的变化。从信息技术的角度看，云计算是工业化的进一步发展，它将有望改变信息技术行业的整体结构。通过使用云计算的服务，软件服务商或开发者不需要自己拥有大量的计算资源（包括计算机服务器硬件和软件）和信息管理人员，就能开发和运行支持多用户的网上软件，或为客户提供托管和使用网上软件的服务。

那么，如何使用云计算服务呢？举例来说，如果想在互联网上建立一个网上售票业务，让人们登录到你的网站后，利用信用卡购票，那么你会首先开发这一网上软件，然后在其他公司的托管服务器和设定空间上运行，或在自己的服务器运行，为顾客服务。这种方式的潜在问题是服务的资源是恒定的，但互联网上的顾客流量是变化的。顾客少时资源可能浪费，顾客多时资源可能不够，顾客用不了软件，影响生意。利用云计算，给顾客提供服务的计算资源可随时根据顾客流量减少或增加，大大提高了服务质量。

1.1.3 云计算的使用范围

除了微软以外，目前提供云计算服务的美国公司还有亚马逊（Amazon）、谷歌（Google）、国际商用机器（IBM）、Saleforce.com 和 EMC，等等。例如，亚马逊提供虚拟的服务器和托管环境及储存空间，但用户必须自己提供或购买服务器和数据库软件；谷歌提供的是编辑语言 Python 和 Java 的托管环境，用户可以将自己的软件送到谷歌云计算环境中运行；IBM 则是为客户提供开发和测试环境；Saleforce.com 为用户提供网上管理客户的软件服务；EMC 提供存储技术服务。

云计算为用户提供的是动态、可扩展的计算资源，也就是说，用户享用的计算资源可以根据客户流量需求随时增减。云计算的特点对现有的企业，特别是对计算资源要求随时间变化的企业具有相当大的吸引力。利用云计算的弹性资源，企业解决了因需求量突然增加而出现计算资源不足的问题，同时避免了因闲置过剩计算资源而造成的浪费。

云计算也特别适合刚刚起步的 IT 企业。新生的企业如果要提供网络服务，通常需要购买一定的服务器等硬件设备和软件，甚至还会招聘管理和支持这些服务器和设备的信息技术管理人员。这对新企业而言是一笔不小的启动资金。利用云计算服务，企业可以花费较少的资金从云计算服务商那里获得所需的网络计算资源，随着业务的发展，再决定是否逐步增加租用云计算服务，甚至设立自己的数据中心。如果企业决定改变经营方向，也不用丢弃现有设备，另起炉

灶，从而风险相对小一些。

随着云计算的普及，人们开发的软件将会越来越多地借助互联网的强大功能，更多的软件将在互联网上直接为用户提供服务，这将给软件开发者（无论企业还是个人）带来他们的黄金时代。如果软件开发者有自己的思想和创意，那么在没有很多经费购买硬件和软件的情况下，借助云计算就有望开发出独特的软件。云计算服务对软件开发方面将起到积极的推动作用，软件的开发也会借此东风向前迈进一大步。

但是，应该指出，不是所有的软件都需要搬到云计算中。云计算也不是对每个开发商都适合。对计算资源需求不大，所需资源没有大起大落的网上软件，从目前来说，云计算并不能带来特别的好处。此外，一些国家和地区，有明确的法律和规章，不容许有关的数据和信息储存在其他国家的数据中心。毫无疑问，云计算在这些国家和地区的使用将受到一定的限制。

1.1.4 云计算与一般托管环境的差别

云计算和一般数据中心的服务器托管听起来很相似，但实际上存在着差别。

首先，工作环境建立有所不同。目前的数据中心提供的托管环境有共享的，也有专用的，有硬件服务器，也有虚拟服务器，但计算资源对于每个托管的软件都是有限的。如果需要更多的资源，就得增加服务器。而云计算的环境可以随时提供所需资源。例如，微软的云计算，开发者不需要和服务器直接打交道，而是与服务模块打交道。为了服务更多的客户，开发者只需指定有多少个软件同时运行。至于数据中心的服务器的启动和管理，由体系管理器来负责。

其次，两者的收费方式也有所不同。服务器托管服务环境通常是按月向用户收取固定费用；云计算服务商则根据计算的时间、信息存储量、计算量等向用户收费。存储量增大，用量增大，信息流量增大，收费也随之增加。

云计算的兴起，对许多公司来说，既是机会，也是挑战。谁能抓住这样的机会，根据市场的需要，提出具有创造力和技术含量的服务，谁就能在竞争中占据胜者之地。

1.2 云产生的背景

有人说云计算是技术革命的产物，也有人说云计算只不过是已有技术的最新包装，是设备厂商或软件厂商新瓶装旧酒的一种商业策略。我们认为，云计算是社会、经济的发展和需求的推动，以及技术进步和商业模式转换共同作业的结果。

1.2.1 经济方面

1. 全球经济一体化

后危机时代加速了全球经济一体化的发展。实践证明：国家和地区的区位优势和比较优势

自发地球寻租，基于成本考虑，价值链的协作者自发整合；基于效率考虑，协同效应需要弹性的业务流程支持。对成本和效率的需求促进云计算的加速发展。

2. 日益复杂的世界和不可确定性的黑天鹅现象

在复杂的世界面前，不确定因素在更快、更广地涌现，计划跟不上变化，任何一台精于预测的机器也无法准确预测到黑天鹅现象的发生（不可预知的未来，一旦发生，影响力极大，事前无法预测，事后有诸多理由解释）。实时的信息获取和全面的信息分析有助于管理复杂性，而按需即用的计算资源、随需应变的业务流程将黑天鹅的负面影响降到最小。实时的、覆盖全网的、随需应变的云计算作业的优势显而易见。

3. 需求是云计算的发展动力

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、使用成本、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，也不需要购买、维护、安放有形的 IT 产品，可以低成本、高效率、随时按需使用 IT 服务；云计算服务提供商可极大地提高资源（硬件、软件、空间、人力能源等）的利用率和业务响应速度，有效聚合产业链。

1.2.2　社会层面

1. 数字一代的崛起

未来的世界在网上，世界的未来在云中。根据埃森哲的调查，中国网民数在 2009 年达到 3.84 亿，超过美国和日本的总和，预计这一数字到 2015 年将增加到 6.5 亿以上。到 2015 年，预计互联网的渗透率将从目前的 29%增加到接近 50%，在中国广大的农村人口中渗透率接近 40%以上。

2. 消费行为的改变

社交网络将现实生活中的人际关系以实名制的方式复制到虚拟世界中，未来网络的发展将是实名制、基于信任和社交化。在线上线下两个世界，半人马型消费者（美国沃顿商学院营销系主任约瑞姆·杰瑞·温德等，《聚合营销——与半人马并驾齐驱》）互相影响，进而影响着为之服务的商业社会和政府行为（如 Dell 基于 Twitter 的营销，广东警方使用微博与民众交流，香港官员使用 Facebook 与民众直接对话）。如图 1-2 所示，历经 10 年的互联网技术和市场的发展，云计算使得数字一代崛起成为可能。

3 亿中国宽带用户中，92%（年龄大于 13 岁）的用户参与到社会化媒体中，而美国仅仅 76%；中国拥有超大规模的社会化媒体的内容贡献者，他们使用博客、微博、社区、视频和图片分享等；43%的中国宽带用户（约 1.05 亿）会使用论坛和 BBS；在中国，25～29 岁的年轻上班族是社交媒体的最活跃用户，与其他年龄段的互联网用户相比，他们更依赖于在线交流的方式；37%的博主（约 2900 万）每天都会更新博客；以一个星期为例，4100 万的中国人是重度的社会化媒体使用者（有 6 个以上的线上活动），会和 54 个人建立联系。云计算是对数字一代消费者提供服务的回答。

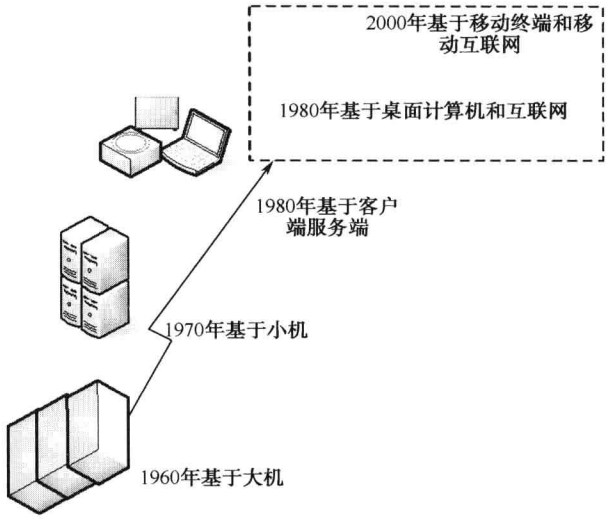


图 1-2　云计算对“数字一代”提供服务效果图

1.2.3　政治层面

1. 社会转型

出口型向内需型社会转型，如何满足人民大众日益增长并不断个性化的需求是一项严峻的挑战。

2. 产业升级

制造型向服务型、创新型的转变。

3. 政策支持

“十二五”规划对物联网、三网融合、移动互联网以及云计算战略的大力支持。

1.2.4　技术方面

1. 技术成熟

技术是云计算发展的基础。首先是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、多租户架构、自动管理与部署；其次是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。

可以将云计算理解为 8 个字：“按需即用、随需应变”，使之实现的各项技术已基本成熟（分布式计算、网络计算、移动计算等）。

2. 企业 IT 的成熟和计算能力过剩

社会需求的膨胀、商业规模的扩大导致企业 IT 按峰值设计，但需求的波动性却事实上使大

量计算资源被闲置。企业内部的资源平衡带来私有云需求，外部的资源协作促进公有云的发展。

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）、一次性买断模式是一个颠覆性的革命。

从纯粹的技术角度看，云计算是很多技术的自然发展、精心优化与组合的结束，是这些技术的集大成者；另外，如果同时考虑到商业模式，那么可以断言，云计算将给整个社会信息化带来革命性的改变。所以，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、无根之木。

1.3 云计算特点

云计算的关键特征如下。

1. 灵活性

灵活性是指一种对资源快速和弹性提供/释放的能力。对消费者来讲，所提供的这种能力是无限的（随需的、大规模的计算机资源），并且在任何时间以任何量化方式可购买的。

2. 成本

云计算可极大降低成本。其主要表现在：基础设施典型地通过第三方提供且不需要一次性购买。

3. 设备和位置独立

消费者无须同服务提供商交互就可以自动地得到自助的计算资源能力，如服务器的时间、网络存储等（资源的自助服务）。

4. 多种租赁

根据消费者的需求来动态地划分或释放不同的物理和虚拟资源，这些池化的供应商计算资源以多租户的模式来提供服务。用户经常并不控制或了解这些资源池的准确划分，但可以知道这些资源池在那个行政区域或数据中心，例如，包括存储、计算处理、内存、网络带宽以及虚拟机个数等。

5. 可靠性

云计算可使企业的业务长久持续，并对发生灾难的文件数据具有恢复性。

6. 可测量性

通过（在要求时）资源的动态规定关于细致的、自助服务基础接近真实时间，不强加最大负荷到用户的工程师。性能被监控及一致的松耦合体系统结构的建造使用 Web 服务作为系统接口。

7. 安全性

典型地改善由于数据集中化，增加关注安全的资源等，但涉及可坚持控制的损失关于特定敏感数据和用户识别的验证。安全通常是与传统系统一样好或更好，部分因为提供方能够贡献资源解决众多客户不能承担的安全问题。提供方典型的登录访问，但访问审计日志本身可能很难或不可能。所有权、控制和访问的数据控制通过“云”提供方可能使得更困难，就像有时很难用目前的效用获取访问“在线”支持。云的范围之下，敏感数据和其他安全相关功能（例如，用户的登记和用户识别的验证）的管理可放置由云提供方和第三方控制。

8. 无所不在的网络访问

云计算可借助于不同的客户端来通过标准的应用对网络访问的可用能力。

1.4 云时代的七大益处

1. 数据集中存储

云时代中的数据集存储的益处主要表现在如下两方面。

1）减少数据泄露

减少数据泄露也是云服务供应商谈论最多的一个。在云计算出现之前，数据常常很容易被泄露，尤其是便携笔记本电脑的失窃成为数据泄露的最大因素之一。为此需要添置额外备份磁碟机，以防数据外泄。随着云技术的不断普及，数据“地雷”也将会大为减少。掌上电脑或者 Netbook 的小量、即时性的数据传输，也远比笔记本电脑批量传输所面临的风险小。询问任何一家大公司的信息安全管理人员（Certified Information Security Office，CISO），是不是所有的笔记本电脑都安装有公司授权的安全技术，如磁盘加密技术（Full Disk Encryption），他们会告诉你这是不大现实的。尽管在资产管理和数据安全上投入了不少精力，但是他们还是面临不少窘境和困难，更何况那些中小企业，那些使用数据加密或者对重要数据分开存储的企业，可以说少之又少。

2）可靠的安全监测

数据集中存储更容易实现安全监测。如果数据被盗，后果不堪想象。通过存储在一个或者若干个数据中心，数据中心的管理者可以对数据进行统一管理，负责资源的分配、负载的均衡、软件的部署、安全的控制，并拥有更可靠的安全实时监测，同时，还可以降低使用者成本。

2. 事件快速反应

云时代中的事件快速反应的益处主要表现在如下几方面。

1）取证准备

在必要的时候，可以利用基础架构即服务（Infrastructure-as-a-Service，IaaS）供应商提供的条件，为自己公司建立一个专门的取证服务器。当事件发生需要取证时，只需要支付在线存储所产生的费用。而不需要额外配置人员去管理远程登录及其软件，而所要做的，只是单击云提供商 Web 界面中的一些按钮。如果一旦产生多个事件反应，可先复制一份，并把这些取证工作