



名师在线丛书

云计算与数据的应用

YUNJISUAN YU SHUJU DE YINGYONG

陈潇潇 王鹏 徐丹丽 ◆ 著

延边大学出版社

云计算与数据的应用

陈潇潇 王鹏 徐丹丽 著

延边大学出版社

图书在版编目 (C I P) 数据

云计算与数据的应用 / 陈潇潇, 王鹏, 徐丹丽著

-- 延吉 : 延边大学出版社, 2018.10

ISBN 978-7-5688-6043-7

I. ①云… II. ①陈… ②王… ③徐… III. ①云计算—数据处理 IV. ①TP393.027②TP274

中国版本图书馆 CIP 数据核字 (2018) 第 246098 号

云计算与数据的应用

著 者: 陈潇潇 王鹏 徐丹丽

责任编辑: 文 熠

封面设计: 刊 易

出版发行: 延边大学出版社

社 址: 吉林省延吉市公园路 977 号 邮 编: 133002

网 址: <http://www.ydcbs.com> E-mail: ydcbs@ydcbs.com

电 话: 0433-2732435 传真: 0433-2732434

制 作: 山东延大兴业文化传媒有限责任公司

印 刷: 延边延大兴业数码印务有限责任公司

开 本: 787×1092 毫米 1/16

印 张: 25.25

字 数: 350 千字

版 次: 2018 年 10 月第 1 版

印 次: 2018 年 10 月第 1 次

书 号: ISBN 978-7-5688-6043-7

定价: 48.00 元

前言

现代社会正以不可想象的速度产生海量数据。人类已经进入数据爆炸性增长的时代——大数据时代。

不断积累的大数据包含着很多在小数据量时不具备的深度知识和价值，带来了巨大的技术创新与商业机遇。大数据分析挖掘将为行业/企业带来巨大的商业价值，实现各种高附加值的增值服务，进一步提升行业/企业的经济效益和社会效益。谈到大数据，不可避免地要提及云计算，云计算结合大数据，这是时代发展的必然趋势。有人把云计算和大数据比作一个硬币的两面。云计算是大数据的 IT 基础和平台，而大数据是云计算范畴内最重要、最关键的应用。大数据必须架构在云上才能高效运作并对外服务，两者缺一不可，相辅相成，相互促进。

大数据云计算正在快速发展，相关技术热点也呈现百花齐放的局面，业界各大厂商纷纷制定相应的战略，新的概念、观点和产品不断涌现。大数据和云计算作为新一代 IT 技术变革的核心，必将成为广大学生、科技工作者构建自身 IT 核心竞争能力的战略机遇。因而作为高层次 IT 人才，学习大数据和云计算知识，掌握相关技术及应用迫在眉睫。

云计算是现代科技飞速发展的成果，在云计算的基础上，人们完善了对于不同高度的大数据分析和应用。本书首先对云计算、大数据进行概述，对云计算与大数据的关系进行了介绍，然后对分布式计算框架、云计算的虚拟化技术、云计算数据安全保障机制一一进行阐述，最后对云计算以及大数据的应用进行了分析。本书集中反映了云计算与数据应用的新思路、新观点、新方法和新成果；注意从实际出发，采用读者容易理解的体系和叙述方法，深入浅出、循序渐进地帮助读者把握云计算与数据应用的主要内容，富有启发性。

本书共计 12 章，合计 35 万字，由湖南文理学院的陈潇潇（负责第一章至第六章，约 13 万字）作者负责部分由湖南省教育厅科学研究项目“云环境下基于 Apriori 算法的水政数据关联规则分析研究”（17C1081）项目资助；内蒙古赤峰学院计算机与信息工程学院的王鹏（负责第七章、第八章，约 11 万字）；上海开放大学的徐丹丽（负责第九章至第十二章，约 11 万字）共同撰写。由于时间比较仓促，加上作者水平有限，在撰写的过程中难免出现纰漏之处，敬请读者批评指正。

目 录

第一章 云计算概述.....	1
第一节 云计算的产生与发展历程.....	1
第二节 云计算名称的由来.....	12
第三节 云计算概念及特点.....	13
第四节 云计算的类型.....	17
第五节 云计算的商业模式.....	21
第二章 大数据概述.....	24
第一节 大数据的产生与发展.....	24
第二节 大数据的概念与特点.....	27
第三节 大数据的技术体系.....	29
第三章 云计算与大数据的关系.....	34
第一节 云计算与大数据的关系研究.....	34
第二节 云计算是大数据处理的基础.....	36
第三节 大数据是云计算的延伸.....	36
第四章 分布式计算框架.....	37
第一节 分布式计算概念.....	37
第二节 Hadoop 系统介绍.....	40
第三节 分布式文件系统.....	42
第四节 MapReduce 计算模式.....	46
第五节 分布式协同控制.....	49
第六节 Spark 计算框架.....	53
第七节 Flink 计算框架.....	60
第五章 云计算的虚拟化技术.....	63
第一节 虚拟化的概述.....	63
第二节 虚拟化分类.....	66
第三节 应用与桌面虚拟化.....	73
第四节 服务器与网络虚拟化.....	77
第五节 存储虚拟化.....	87

第六节 本章小结.....	90
第六章 云计算平台.....	92
第一节 Google 云计算平台.....	92
第二节 Amazon 云计算平台.....	97
第三节 Microsoft 云计算平台.....	106
第四节 阿里云计算平台.....	108
第五节 开源云计算平台.....	116
第六节 云计算仿真平台.....	143
第七章 云计算大数据存储.....	146
第一节 非结构化数据分析.....	146
第二节 NoSQL 数据库系统.....	160
第三节 分析型数据库系统.....	188
第八章 云计算绿色数据中心.....	210
第一节 绿色云计算.....	210
第二节 云计算绿色数据中心概述.....	220
第三节 云计算中的大数据分析.....	222
第四节 云环境下数据存储优化.....	233
第五节 数据中心的管理与维护.....	244
第六节 数据聚集算法与实验分析.....	246
第九章 云计算大数据面临的安全威胁.....	255
第一节 云计算大数据的安全问题.....	255
第二节 大数据在不同领域的安全需求.....	262
第三节 大数据的安全内涵.....	265
第四节 大数据安全研究方向.....	266
第十章 云计算数据安全保障机制.....	270
第一节 云计算的安全分析.....	270
第二节 云数据销毁.....	276
第三节 云存储数据的隐私保护.....	307
第四节 大数据安全的关键技术.....	332
第五节 大数据安全保障实践.....	340
第十一章 云计算的应用.....	346
第一节 在电信领域的应用.....	346

第二节 在医疗领域的应用.....	352
第三节 在政务领域的应用.....	356
第四节 在电子商务领域的应用.....	362
第十二章 大数据的应用.....	367
第一节 在媒体领域的应用.....	367
第二节 在医疗领域的应用.....	371
第三节 在金融领域的应用.....	374
第四节 在电子商务领域的应用.....	386
第五节 在教育领域的应用.....	389
参考文献.....	393

第一章 云计算概述

第一节 云计算的产生与发展历程

20 世纪，图灵奖得主、美国计算机科学家、认知科学家 John McCarthy 提出“计算迟早有一天会变成一种公用基础设施”的设想。计算技术、网络等信息技术的高速发展，尤其是近年出现的云计算（Cloud Computing）的技术和理念，将 John McCarthy 的设想演变为现实。作为一种新兴的信息服务模式，云计算已经深入各个行业，并带来了巨大的效益。事实上，云计算已经深入我们普通人生活的方方面面，不管我们是否意识到，我们已经离不开云计算了。

本章主要向读者全面介绍云计算的产生与发展：首先介绍云计算的产生背景，然后描述传统计算模式到云计算模式的发展历程，最后探讨云计算的定义典型特征，以及与其他计算模式的区别。

一、云计算的产生

在传统模式下，企业建立一套信息服务系统不仅需要购买硬件等基础设施，还要买软件的许可证，并需要专门人员维护。当企业的规模扩大时又要继续升级各种软、硬件设施以满足需求。对于企业来说，计算机硬件和软件本身只是他们完成工作、提供效率的工具而已，无须独占拥有。对个人来说，使用计算机也需要安装许多软件，而许多软件并不经常使用，常常处于闲置状态，对用户来说这显然浪费。能不能有这样一种服务或平台，给我们提供可以动态租用的软、硬件资源？云计算正是为满足这种需求而诞生的。

云计算的想法可以追溯到 20 世纪 60 年代，John McCarthy 曾经提到“计算迟早有一天会变成一种公用基础设施”，即计算能力可以像煤气、水电一样，取用方便、费用低廉。云计算最大的不同在于它提供的服务和资源是通过 Internet 进行传输的。从最根本的意义来说，云计算就是数据、应用和服务均存储在云服务器端，充分利用云数据中心（Cloud Data Centre）所拥有的规模庞大的服务器集群（Cluster）的强大计算能力和海量存储资源，实现用户业务系统的自适应性部署和高效运行。2007 年 10 月，IBM 和 Google 宣布在云计算领域

进行合作，以及 Google 发布了一系列云计算技术论文，使云计算吸引了众人的关注，并迅速成为产业界和学术界研究的热点。

21 世纪初，Web 2.0 的流行让网络迎来了新的发展高峰。网络服务系统需要处理的业务量快速增长，例如，在线视频或照片共享网站、社交网络平台需要为用户储存和处理大量的数据，这类系统所面临的重要问题是：如何在用户及服务数量快速增长的情况下快速扩展原有系统？随着移动终端的智能化和移动宽带网络的普及，越来越多的移动终端设备进入 Internet，这意味着与移动终端相关的信息系统会承受更多的负载，而对于提供数据服务的企业来讲，其信息系统需要处理更多的业务量。由于资源的有限性，其电力成本、空间成本、各种设施的维护成本快速上升，直接导致数据中心的成本上升，这就面临着如何有效地、更少地利用资源处理更多任务的问题。同时，处理器芯片和存储设备在性能增强的同时，价格也在变得更加低廉，拥有大规模服务器集群的数据中心，也具备了快速为大量用户处理复杂问题的能力。

技术上，分布式计算（Distributed Computing）技术的日益成熟和应用，特别是网络计算（Network Computing）的发展通过 Internet 把分散在各处的硬件、软件、信息资源连接成为一个巨大的整体，使得人们能够利用地理上分散于各处的资源，完成大规模、复杂的计算和数据处理任务。数据存储的快速增长产生了以谷歌文件系统（Google File System, GFS）、存储域网络（Storage Area Network, SAN）为代表的高性能存储技术。另外，服务器整合需求推动了虚拟化（Virtualization）技术的进步。此外，多核技术的广泛应用，也为构建更强大的计算能力和服务平台提供了可能。随着对计算能力、资源利用效率、资源集中化的迫切需求，云计算应运而生。

二、云计算发展历程

（一）计算模式演进

云计算是在并行计算（Parallel Computing）、分布式计算、网格计算（Grid Computing）和效用计算（Utility Computing）的基础上发展起来的，经过持续演化和融合改进逐步形成目前流行的云计算模型，云计算的演化过程如下所述：

1. 并行计算

在单核多线程的系统设计中，采用的算法均基于串行计算模式，即将任务分解成一串相互独立的命令执行流，每个命令执行流有自己的序号，串行计算要求所有的命令执行流按照顺序逐一执行，也就是说同一时间只有一个执行流

在执行。

这种算法效率低下，无法满足大数据（Big Data）的分析和处理需求，而并行计算可以通过同时调用多个计算资源处理庞大、复杂的计算任务。这些计算资源可以是拥有多核 CPU 或者多个 CPU 的高性能服务器，也可以是多台服务器组成的集群系统。

任务可以分解成相互独立却可以同时运行的部分，每一部分再分解成相互独立的命令执行流。任务分解后，每部分的每个命令执行流都可以同时得以执行。

并行计算包括空间并行、基于流水线（Pipeline）技术的时间并行，以及基于优化算法的数据并行和任务并行等。不管采用何种并行计算方法，都对串行计算的单指令流单数据流（Single Instruction Single Data, SISD）做出优化，以及通过采用多指令流多数据流（Multiple Instruction Stream Multiple Data Stream, MIMD）的并行计算大幅度提升系统的处理能力。

2. 分布式计算

并行计算调动的计算资源可以是单个 CPU 的多个内核或单个服务器内的多个 CPU，也可以是服务器集群提供的多个 CPU。如果仅从这个角度上看，分布式计算和并行计算有相似之处。

MapReduce 等分布式计算模式在处理庞大的计算请求时，会将需要解决的问题分解成细小的组成部分，然后将这些组成部分散给众多的计算机进行处理，处理完成后将结果进行汇总，形成最终结果。分布式计算则可以汇集成千上万台计算机，甚至几百万、几千万的计算机资源。

分布式计算的典型代表是对等计算（Peer-to-Peer Computing, P2P）。P2P 使得 Internet 用户可以提供其个人计算机上闲置的处理能力和存储资源，通过资源共享和计算能力的平衡负载来提供分布计算服务。目前，众多机构发起了不同的 P2P 分布式计算项目，以解决复杂的数据难题、密码分析、生物科学、数据处理等大规模计算问题，如利用全球联网的计算机共同搜寻地外文明的 SETI@Home 项目、为大型强子计算机提供计算能力的 LHC@home 项目等。有兴趣的读者也可以关注伯克利开放式网络计算平台（Berkeley Open Infrastructure for Network Computing, BOINC）并搜寻自己感兴趣的项目。

3. 网格计算

根据 Larry Smarr 的描述，网格计算系统是一种无缝、集成的计算和协作环境。按照网格提供的功能，网格可分为两类：计算网格（Computational Grid）

和存储网格（Access Grid）。计算网格可以提供虚拟的、无限制的计算和分布数据资源，存储网格则提供一个合作环境。

网格计算系统一般具有如下特点：

（1）异构性：网格可以包含多种异构资源，包括跨越地理分布的多个管理域。构成网格计算系统的超级计算机有多种类型，不同类型的超级计算机在体系结构、操作系统及应用软件等多个层次上可能具有不同的结构。

（2）可扩展性：网格可以从最初包含少数的资源发展到具有成千上万资源的大网格。

（3）可适应性：网格中具有很多资源，资源发生故障的概率很高。网格的资源管理或应用必须能动态适应这些情况，调用网格中可用的资源和服务来取得最大的性能。

（4）不可预测性：在网格计算系统中，由于资源的共享造成系统行为和系统性能经常变化。

（5）多级管理域：由于构成网格计算系统的超级计算机资源通常属于不同的机构或组织并且使用不同的安全机制，因此需要各个机构或组织共同参与解决多级管理域的问题。

相比于分布式计算来说，网格计算不仅仅是一种计算模式，更是一套广泛的整合各种异构计算资源方式的方案和思想。

网格计算系统的主要部件包括：

（1）网格基础设施：包含网上可访问的所有资源，如运行 Windows、Linux 或 UNIX 的 PC 或工作站、运行 Cluster 操作系统的机群、存储设备、数据库，也可能是科学仪器。

（2）网格中间件（Grid Middleware）：网格中间件提供核心服务，如远程进程管理服务、资源分配服务、存储访问服务、信息服务、安全控制服务、质量服务。

（3）网格发展环境和工具：网格必须提供网格应用开发工具。

（4）网格应用和网格门户：可以使用 PVM、MPI 等工具开发参数模拟等应用，这些应用通常需要相当多的计算资源以及远程数据访问。网格门户提供基于 Web 的应用服务，用户通过网络界面提交任务，并得到结果。

4. 效用计算

为了解决传统计算机资源、网络及应用程序的使用方法变得越来越复杂，并且管理成本越来越高的问题，科学家们提出了“效用计算”这个概念。效用

计算的具体目标是结合分散各地的服务器、存储系统及应用程序来立即提供需求数据的技术，使得用户能够像把电器插头插入插座取电一样来使用计算机资源。“效用”（Utility）这个词是指为客户提供个性化的服务，并且可以满足不断变化的客户需求，可以基于实际占用的资源进行收费。

按需分配的效用计算模型采用了多种灵活有效的技术，能够对不同的需求提供相应的配置与执行方案。效用计算使用户可以通过网络来连接资源并实现企业数据的处理、存储和应用，而企业不必再组建自己的数据中心。

效用计算模型中包括计算资源、存储资源、基础设施等众多资源，它的收费方式发生了改变，不仅仅对速率进行收费，对于租用的服务也需要缴纳一定的费用。这种按照实际使用进行收费计费方式在企业中变得越来越常见。从效用计算开始引入“按需计算”（Computing on Demand）的理念，不需要的额外服务不必为其支付任何费用。它的管理模块注重系统的性能，确保数据和资源随时可用，同时建立自动化（Automatization）模块，对服务器进行集群操控，促进服务器之间的自动化管理，保证服务之间可以自行分配。

可以看出，效用计算已经开始有了很多云计算的影子，云计算的很多理念也是在效用计算的基础上发展起来的。

5.云计算

云计算强调所有资源均以服务的形态出现，包括基础设施即服务（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）、数据即服务（Data as a Service, DaaS）、知识即服务（Knowledge as a Service, KaaS）、存储即服务（Storage as a Service, SaaS）、安全即服务（Security as a Service, SECaaS）等。

企业对信息中心提出的要求会越来越高，企业首席信息官（Chief Information Officer, CIO）更加希望从基础设施、平台、软件中摆脱出来，转而关注业务流程的革新、办公效率的优化、业务成本的管控，这会给信息系统的交付和管理模式带来变化。中小企业则希望避免自行构建数据中心，而是将所有的服务迁移到公有云（Public Cloud）；大型企业则可以建立私有云（Private Cloud），将所有的资源整合，再以服务的形态呈献给企业内部员工和外部用户。而对于个人用户来说，理想的状态显然是通过一台能联网的设备来完成所有的办公、生活和娱乐需求，而不管身处何方，也不管使用的是笔记本电脑还是智能手机等移动终端。

简而言之，云计算通过各种不同计算模式不断地演变、优化，形成我们现

在所看到的云计算，它的发展不仅顺应当前计算模型，也为企业真正地带来效率和成本方面的诸多变革。

（二）云计算发展大事记

云计算被认为是科技界的一次巨大革新，已经带来工作方式和商业模式的根本性改变。几十年来，云计算发展历程的里程碑事件列举如下：

1959 年，Christopher Strachey 发表论文提出虚拟化概念，虚拟化目前已成为云计算的核心技术之一。

1961 年，John McCarthy 提出计算力和通过公用事业销售计算机应用的思想。

1962 年，Joseph Carl Robnett Licklider 提出“星际计算机网络（Intergalactic Computer Network）”设想并详细阐述了这一概念，“星际计算网络”的概念包含了现代因特网几乎所有的特征。

1965 年，美国电话公司 Western Union 提出建立信息公用事业的设想。

1984 年，Sun 公司的联合创始人 John Gage 指出“网络就是计算机”，今天的云计算正在实践这一理念。

1996 年，开源网格计算平台 Globus 被推出，Globus 是基于开放结构、开放服务资源和软件库，并支持网格和网格应用，目的是为构建网格应用提供中间件服务和程序库。

1997 年，南加州大学教授 Ramnath K. Chellappa 首次提出云计算的学术定义，认为计算的边界可以不是技术局限，而是经济合理性（Computing paradigm where the boundaries of computing will be determined by economic rationale rather than technical limits alone.）。

1998 年，VMware 公司成立并首次引入 x86 的虚拟化技术，即基于 x86 处理器架构的虚拟化技术，虚拟化之后可以在同一台服务器硬件上运行多个 Guest OS（Guest OS 是指虚拟机所用的操作系统）。

1999 年，Marc Andreessen 创建 LoudCloud，这是第一个商业化的 IaaS 平台。1999 年，Salesforce 公司成立，并以“软件终结（NO SOFTWARE）”为口号，目前 Salesforce 已成为全球最大的网络商用管理软件销售商。

2000 年，软件即服务（SaaS）开始被广泛接受，这意味着 SaaS 的兴起。

2004 年，在出版社经营者 O'Reilly 和 MediaLive International 之间的一场头脑风暴论坛中，Web 2.0 的概念诞生，自此以后 Web 2.0 成为技术流行词，Internet 发展进入新阶段。同年，Google 发布 MapReduce 论文，Google 在这篇论文中阐述了公司内部已经广泛采用的基于 Map 过程和 Reduce 过程的海量数据处理

思想，这篇论文的发布在大数据兴起的历程中意义重大。Doug Cutting 和 Mike Cafarella 实现了 Hadoop 分布式文件系统（HDFS）和 Map-Reduce，Hadoop 目前已成为了非常优秀的分布式系统基础架构，Hadoop 主要由 HDFS、MapReduce 和 Hbase 组成。

2005 年，Amazon 宣布 Amazon Web Services（AWS）云计算平台。

2006 年，Amazon 相继推出在线存储服务（Simple Storage Service, S3）和弹性计算云（Elastic Compute Cloud, EC2）等云服务。同年，Sun 推出基于云计算理论的“BlackBox”计划。

2007 年，Google 与 IBM 在大学开设云计算课程。同年，Dell 成立数据中心解决方案部门，先后为全球著名云计算平台 Windows Azure、Facebook 和 Ask.com 提供云计算基础设施。亚马逊公司推出了简单队列服务（Simple Queue Service, SQS），通过这项服务，开发人员可以开发分布式的应用程序，并在它们中间以一种安全、灵活和可靠的方式进行通信。IBM 首次发布云计算商业解决方案，推出“蓝云（Blue Cloud）”计划。

2008 年，Salesforce 推出了按需应变平台 DevForce，这是世界上第一个 PaaS 应用。同年，IBM 宣布在无锡太湖新城科教产业园为中国软件公司建立云计算中心。Google 发布 PaaS 平台 Google App Engine，该平台可以让开发者在 Google 提供的基础架构上运行自己的网络应用程序。Gartner 发布报告，指出云计算代表了计算的方向。Sun 在 JavaOne 开发者大会上宣布推出“Hydrazine”计划，在该计划中，Sun 计划利用其大多数核心技术提供一个功能齐全的解决方案，让开发人员可以利用 Sun 的平台创建应用程序和服务，并且可以利用这些应用程序和服务赚钱。EMC 公司中国研发中心启动“道里（Daoli）”可信基础架构联合研究项目。IBM 宣布成立 IBM 大中华区云计算中心。HP、Intel 和 Yahoo 联合创建云计算试验台 OpenCirrus。美国专利商标局网站信息显示，Dell 申请云计算商标，此举旨在加强对这一未来可能重塑技术架构的术语的控制权，Dell 在申请文件中称，云计算是“在数据中心和巨型规模的计算环境中，为他人定制制造计算机硬件”。Google 公司推出 Chrome 浏览器，将浏览器彻底融入云计算时代。Oracle 和 Amazon 云平台（Amazon Web Services, AWS）合作，用户可在云中部署 Oracle 软件以及在云中备份 Oracle 数据库。Citrix 公布其云计算战略，并发布 Citrix 云中心（Citrix Cloud Center, C3）产品系列。Microsoft 发布其公开云计算平台 Windows Azure，由此开启了 Microsoft 的云计算战略。Gartner 指出十大数据中心关键性技术，虚拟化和云计算上榜。Amazon、Google

和 Flexiscale 的云服务相继发生宕机故障，引发业界对云计算安全的讨论。

2009 年，Cisco 发布其云计算服务平台——统一计算系统（UCS），并与 EMC、VMware 联合建立虚拟计算环境联盟，这是三大 IT 业界领袖企业的首次共同协作。该联盟的成立，旨在通过普适数据中心虚拟化和向私有云架构的转型，不断提高 IT 基础架构的灵活性，降低 IT、能源和空间成本，从而让客户能够快速提高业务敏捷性。同年，Spark 大数据计算平台诞生于伯克利大学 AMPLab 实验室，并于 2010 年开源，2013 年 Spark 正式成为了 Apache 基金项目。阿里软件在江苏南京建立首个电子商务云计算中心。VMware 推出业界首款云操作系统 VMware vSphere。Google 宣布将推出操作系统 Chrome OS。中国首个企业云计算平台——中化企业云计算平台诞生。VMware 启动 vCloud 计划，构建全新云服务。中国移动的云计算计划——“大云（Big Cloud）”项目启动。

2010 年，HP 和 Microsoft 联合提供云计算解决方案。IBM 与松下达成当时全球最大的云计算交易。Microsoft 正式发布 Microsoft Azure 云平台服务。英特尔在英特尔开发者论坛（Intel Developer Forum，IDF）上提出互联计算，通过 X86 架构统一嵌入式计算、物联网和云计算领域。Microsoft 宣布其 90% 员工将从事云计算相关研发工作。Dell 推出源于 DCS（Distributed Control System）部门设计的 PowerEdgeC 系列云计算服务器及相关服务。

2011 年，Amazon 位于美国弗吉尼亚州的云计算数据中心宕机，导致回答服务 Quora、新闻服务 Reddit 和 Hootsuite、位置跟踪服务 FourSquare 及为网络出版商提供游戏工具的 BigDoor 瘫痪，故障持续了 4 天，被认为亚马逊史上最严重的云计算安全事件。开放式数据中心联盟（ODCA）发布了云计算应用模型路线图，涉及提升安全性、服务透明性、计算性能以及互操作性等一系列建议，这些建议将指导联盟内 300 多家会员的采购决策。Apple 发布在线数据存储服务 iCloud，能够将用户在 Mac、iPad 和 iPhone 上的文件自动存储到 Apple 的个人服务器上，用户可在不同介质里同步分享自己的文件。Microsoft 的 BPOS 云托管套件服务再次中断 3 小时，在北美的 Microsoft 用户都受到了影响。国家发改委下拨 7 亿元专项资金支持北京、上海、深圳、杭州和无锡开展 15 个云计算示范项目，获得这批扶持资金的企业包括联想、百度、腾讯、阿里巴巴、华胜天成和金蝶软件等。Oracle 首席执行官拉里·埃里森（Larry Ellison）宣布将推出 Oracle 公有云（Oracle Public Cloud）服务，以及 Oracle 社交网络（Oracle Social Network）平台。Salesforce 宣布收购云平台社交管理公司 Ryppele，Ryppele

拥有包括 Facebook、网络音乐服务提供商 Spotify 在内的 350 个客户。

2012 年，百度召开了百度开发者大会，在大会上发布了其在云计算领域战略规划，并推出了开发者中心及四大服务体系，包括开发支持、运营支持、渠道推广以及商业变现。IBM 宣布加入 OpenStack 项目，并作为主要赞助商。我国科学技术部公布《中国云科技发展“十二五”专项规划》，以加快推进云计算技术创新和产业发展。Viacloud 互联云平台引入 OpenStack 项目，研制基于 OpenStack 的公有云和私有云平台。

2013 年，阿里巴巴集团宣布，旗下的阿里云与万网将合并为新的阿里云计算公司，IBM 在 IBM Pulse 大会上宣布将基于 OpenStack 提供私有云服务及相关应用。2013 年，Microsoft 在上海宣布 Azure 云服务即将进入中国。IBM 高调宣布与首都在线科技股份有限公司签署公有云长期战略合作协议。为成长型企业提供基础云服务的 Ucloud 获风险投资公司 DCM、贝塔斯曼集团的投资，超 1000 万美元。Amazon 宣布即将在中国推出云计算平台服务。

2014 年，Microsoft 宣布 Microsoft Azure 在中国正式商用，用户熟悉的 Office 套件、即时通信、协作组件将首次作为云服务提供给中国企业用户及政府机构，由世纪互联负责运营。可信云服务大会公布了第一批通过“可信云服务认证”的名单，包括中国电信、中国移动、BAT、华为、京东、世纪互联等 19 家云服务商的 35 项云服务。阿里云启动“云计划”，该计划拟招募 1 万家云服务商，为企业、政府等用户提供一站式云服务。小米 CEO 兼金山软件董事长雷军，宣布金山软件未来 3~5 年内将会向云业务进行规模超过 10 亿美元的投资，而金山云未来 3 年将执行“All in Cloud”的战略。

2015 年，云栖大会以“Internet、创新、创业”为本届主题，展现“互联网+”时代无处不在的云计算与各行各业的连接。

2016 年，Microsoft 宣布正式开放 Azure Container 服务。AWS 技术峰会在芝加哥举办，Amazon 宣布 AWS Snowball 能够以安全可靠的方式将企业的大量数据传输进入 AWS 或者从 AWS 提取出来。腾讯的“云+未来峰会”上马化腾宣布腾讯云开放生态体系资源。百度李彦宏在“2016 百度云计算战略发布会”宣布百度的“云计算+大数据+人工智能”的“三位一体”战略。Google 收购云市场平台创业公司 Orbitera，增强其在云计算服务领域与 Amazon、Microsoft 和 Salesforce 竞争的實力。

2017 年，在 VMworld@2017 大会上，VMware 和 Amazon Web Services 共同宣布 VMware Cloud™ on AWS 初步可用。同年，AWS 宣布已经创建了新的基

于 KVM 虚拟化引擎，新的 C5 实例和未来的虚拟机将不使用 XEN，而是核心的 KVM 技术。

2018 年，Redhat 宣布收购 Kubernetes 与容器原生解决方案创新者与领导厂商 CoreOS，交易金额 2.5 亿美金。Redhat 表示：“该收购将进一步增强利用开源的灵活性，让企业用户在任何环境构建并部署任何应用的愿景。该交易将夯实红帽作为混合云与现代应用开发基石的地位。”

（三）云计算时代

云计算引发了新的技术变革和带来了新的 IT 服务模式，目前已成为 IT 领域最令人关注的话题之一，也是各行各业正在考虑和投入的重要领域。

企业的云化 IT 设施建设过程可以分为三个阶段。

第一个阶段：集中化

这一阶段将企业分散的计算、存储与数据等资源进行了集中，形成了规模化的数据中心基础设施。在数据集中过程中，不断实施数据和业务的整合，大多数企业的数据中心基本完成了自身的标准化，使得既有业务的扩展和新业务的部署能够规划、可控，并以企业标准进行 IT 业务的实施，解决了数据业务分散时期的混乱无序问题。在这一阶段中，很多企业在数据集中后期也开始了容灾建设，特别是金融行业的企业大部分建设高级别的容灾系统，以数据零丢失为目标。总体来说，第一阶段过程解决了企业 IT 资源分散管理和容灾的问题。

第二个阶段：虚拟化

在数据集中与容灾实现之后，随着企业的快速发展，数据中心 IT 基础设施扩张很快，但是系统建设成本高、周期长，即使是标准化的业务模块建设（如系统的复制性建设），软硬件采购成本、调试运行成本与业务实现周期并没有显著下降。标准化并没有给系统带来灵活性，集中的大规模 IT 基础设施出现了系统利用率不足的问题，不同的系统运行在独占的硬件资源中，效率低下而数据中心的能耗、空间问题逐步突显出来。因此，以降低成本、提升 IT 系统运行灵活性、提升资源利用率为目的的虚拟化机制开始在数据中心进行应用。虚拟化屏蔽了物理设备的异构性，将基于标准化接口的物理资源虚拟化成逻辑上完全标准化、一致化的逻辑计算资源和逻辑存储空间。虚拟化可以将多台物理服务器整合成单台，每台服务器上运行多种应用的虚拟机（Virtual Machine, VM），实现物理服务器资源利用率的提升。由于虚拟化环境可以实现计算与存储资源的逻辑化变更，特别是虚拟机的克隆，使得数据中心 IT 系统部署的灵活性大幅提升，业务部署周期由数月缩小到一天以内。虚拟化后，应用以 VM 为单元部