“十二五”职业教育国家规划教材
经全国职业教育教材审定委员会审定

21世纪高等院校
云计算和大数据人才培养规划教材

CLOUD COMPUTING AND BIG DATA

云计算和大数据技术

概念 应用与实战 第2版

王鹏 李俊杰 谢志明 石慧 黄焱 ◎ 编著

- 紧扣实验环节，深入浅出，以任务驱动模式组织内容
- 理论精简，案例经典，突出实操实训
- 配备课件、操作视频、软件资源等电子资源

中国工信出版集团
人民邮电出版社
POSTS & TELECOM PRESS

“十二五”职业教育国家规划教材
经全国职业教育教材审定委员会审定

21世纪高等院校
云计算和大数据人才培养规划教材

CLOUD COMPUTING AND BIG DATA

云计算和大数据技术

概念 应用与实战 第2版

王鹏 李俊杰 谢志明 石慧 黄焱 ◎ 编著

人民邮电出版社
北京

图书在版编目（CIP）数据

云计算和大数据技术 ： 概念、应用与实战 / 王鹏等编著. -- 2版. -- 北京 ： 人民邮电出版社, 2016.8
21世纪高等院校云计算和大数据人才培养规划教材
ISBN 978-7-115-42080-0

Ⅰ. ①云… Ⅱ. ①王… Ⅲ. ①计算机网络－数据处理－高等学校－教材 Ⅳ. ①TP393

中国版本图书馆CIP数据核字(2016)第105738号

内 容 提 要

本书全面介绍云计算与大数据的基础知识、主要技术、基于集群技术的资源整合型云计算技术和基于虚拟化技术的资源切分型云计算技术。全书共 10 章，主要内容包括云计算基础与大数据基础、虚拟化技术和平台、MPI、Hadoop、HBase、Hive、Storm 和云存储系统 Swift。本书以“实践为主、理论够用”为原则，注重实用，实验丰富，将实验内容融合在课程内容中，使理论紧密联系实际。

本书主要面向高等院校计算机专业的学生，也可作为其他相关专业云计算、大数据相关课程的教材，以及 IT 类培训机构云计算与大数据等相关课程的培训教材和从事相关技术人员的参考书。

◆ 编　　著　王　鹏　李俊杰　谢志明　石　慧　黄　焱
责任编辑　马小霞
责任印制　焦志炜

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路 11 号
邮编　100164　　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
大厂聚鑫印刷有限责任公司印刷

◆ 开本：787×1092　1/16
印张：12.5　　2016 年 8 月第 2 版
字数：307 千字　　2016 年 8 月河北第 1 次印刷

定价：35.00 元

读者服务热线：(010)81055256　印装质量热线：(010)81055316
反盗版热线：(010)81055315

前　言

本书2013年8月被教育部职业教育与成人教育司批准为“十二五”职业教育国家规划教材（教职成司函〔2013〕184号），第1版于2014年5月出版。由于云计算技术和大数据技术发展迅猛、日新月异，尤其是在2015年10月教育部对普通高等学校高等职业教育（专科）专业目录进行修订的教职成〔2015〕10号文件中明确将云计算技术与应用作为一个专业列入高职高专专业目录中。本书第2版增加了大约30%的篇幅，并按照院校的教学与学习习惯对章节进行了全新的编排，对原有的内容进行全面改写及扩充，以确保能更好地反映云计算专业今后的学习和发展的方向。

为使第2版能更加贴近院校教学需求，本书在改版时不仅对读者进行了调查，还调研了使用本书第1版的多所高职高专院校。反馈回来的结果表明大家更希望的是“理论要精简，案例要经典，实验要步骤明确、易于实操”。因此，本书在改版时采用“理论够用为度，突出实操实训环节”，删除了面向数据的高性能计算集群系统（HPCC）、服务器与数据中心、云计算大数据仿真技术章节，将集群系统基础中的部分内容调整到新版的第1章和第2章；增补了XenServer、VMware vSphere虚拟化平台的安装与部署、HBase、Hive和Swift的实操实训内容，并对MPI、Hadoop、Storm做了较大篇幅的调整。本书紧扣实验环节，深入浅出，以任务驱动模式组织内容，让读者知其然并知其所以然。

本书由西南民族大学计算机科学与技术学院王鹏教授组织编写，是2015年广东省高等职业教育质量工程教育教学改革项目（课题编号：GDJG2015245）和高职教育信息技术教指委教改项目（课题编号：XXJZW2015002）、2016年广东省高等教育学会高职高专云计算与大数据专业委员会教育科研课题（课题编号：GDYJSKT16-01、GDYJSKT16-03、GDYJSKT16-05）、汕尾职业技术学院2014年度资源精品共享课《云计算技术》（课题编号：swzyjpkc14002）、汕尾职业技术学院教学改革与科研立项课题（课题编号：SWKT15-002、SWKT16-002、swjy15-004、swjy15-016）、广州市教育科学规划课题（项目编号：1201420456）、模式识别与智能信息处理四川省高校重点实验室开放基金（课题编号：MSSB-2015-9）和成都市科技局创新发展战略研究项目（项目编号：11RKYB016ZF）的科研成果，本书还得到了广东省高职高专云计算与大数据专业委员会、西南民族大学、广州五舟科技股份有限公司、汕尾市创新工业设计研究院、淮阴师范学院的鼎力支持，同时也得到汕尾职业技术学院各处系领导、老师的支持和帮助，因为有了他们的支持和帮助，我们才能完成本书的撰写和出版工作。

云计算与大数据技术涉及面很广，在第2版的编写过程中部分章节及内容仍然继承了部分第1版编写工作者的经验与成果，同时，还参考并引用了大量前辈学者的研究成果和论述，对此编者向这些学者一并表示深深的敬意。云计算与大数据技术是一门高速发展的技术领域，新技术、新方法、新架构层出不穷，由于作者的经验和能力所限，本书的结构、内容肯定存在许多疏漏和不妥，望读者指正。

为方便读者学习、满足教学需要，本教材配备了大量的电子资源，欢迎读者登录人

民邮电出版社教育服务与资源下载社区（http://www.ryjiaoyu.com）下载或登录并行计算实验室网站（http://www.qhoa.org）免费下载使用，同时还欢迎相关课程的教师加入云计算大数据 HPC 教育 QQ 群(321168742)讨论交流。读者还可以通过发送邮件给编者以获得更多资源（百度云盘和 360 云盘链接及提取码）。编者的 E-mail 是：gdswyun@126.com。

感谢您使用本书，期待本书能成为您的良师益友，也欢迎使用配套教材《云计算和大数据技术实战》（ISBN 978-7-115-39079-0）。

编　者

2016 年 3 月

第1版前言

计算技术的发展经历了从合到分，又从分到合的历程，这一发展历程中内在的推动力就是技术。最早的电子管计算机系统价格昂贵、体积巨大，计算资源只能被集中放在机房，随着芯片技术的发展，大规模集成电路技术使计算机的体积变得很小，同时微软视窗系统的出现使计算以前所未有的速度得到了普及，计算实现了由合到分的变化。分散的计算资源虽然给大家带来了方便，但同时也带来了资源的浪费，而且在需要进行计算时又可能会出现资源不够的情况，这时网络技术的发展使计算资源再次被集中存放于机房成为了可能。

计算技术的发展特别是网络技术的发展催生了云计算技术的出现，云计算技术的出现被广泛地认为是信息技术的一次重大变革，大量的与云计算相关的软件和系统架构如雨后春笋般出现。云计算技术将计算资源、存储资源以及相关各类广义的资源通过网络以服务的形式提供给资源的使用者，改变了传统信息技术架构中物理资源直接独占使用的模式，甚至从广义上讲只要是通过网络向用户提供服务的信息系统都被称为云计算系统。

云计算、物联网、社交网络的发展使人类社会的数据产生方式发生了变化，社会数据的规模正在以前所未有的速度增长，数据的种类五花八门，对海量、异构数据的存储、管理、分析和挖掘成为信息学科的热门领域，大数据技术逐渐进入人们的视野。

云计算与大数据出现以后，随着进入这个领域的企业和研究机构的大量增加，对于云计算、大数据技术的认识出现了大量不同的定义。如果我们把云计算看作是一种通过网络实现资源服务的模式的话，则云计算技术可以被认为是实现云计算模式的所有技术的总称，这些技术包括虚拟化技术、分布式计算技术、分布式存储技术、网络技术等，不少技术是互联网时代就已存在的技术。大数据技术涵盖数据的存储、管理、分析和挖掘，这些技术并不是新的技术门类。

在云计算与大数据概念的内涵还没有完全得到业界的一致认识时，云计算与大数据产业的高速发展却十分出人意料，大量的客户需要企业提供相关的系统解决方案，一些地方希望能建设云计算中心，云计算与大数据人才的需求呈现出一种井喷的局面，不少学校都在规划建立云计算与大数据专业或开设相关课程以满足日益增长的人才需求。云计算与大数据课程专业该“学什么、如何学”正是本教材需要回答的问题。

信息技术这些年的高速发展使信息学科的整个格局也发生了变化。例如，在高性能计算领域已存在很久的集群技术在云计算和大数据时代再次成为系统架构的核心技术；在传统数据中心主机租赁业务中得到广泛应用的服务器虚拟技术，在云计算时代因为桌面虚拟化的大量使用得到了极大的发展，成为云计算技术的重要应用之一。

本书作为云计算与大数据技术的一本综合入门课程，我们一直在思考什么样的人才可以被称为云计算与大数据人才，培养的学生的知识结构是怎么样的，云计算与大数据作为一个高速发展的学科哪些知识是必须要了解的。从课程角度本书并不是对某一项技术的专门介绍，而是希望为学习云计算与大数据技术的同学提供一个完整的知识框架，

为今后深入学习打下基础。

本书主要包含两大技术方向：集群计算技术和虚拟化技术，分别介绍了两个技术方向中学生需要了解的基础知识和典型系统，使学生在面对技术的快速发展时能以不变应万变，避免出现在学校学的技术到工作岗位上由于技术进步而用不上的问题。书中所介绍的相关知识和技术都带有一定的普遍性和典型示范作用，在学习时重要的是要学习其中的系统思想。特别是我们在集群云计算系统中加入了基于消息传递机制高性能计算内容，消息传递机制揭示了集群系统中节点间协调工作和数据传输模式的本质，不少人在学习云计算与大数据技术时知其然而不知其所以然的原因就在于不了解集群工作的基本机制。基于消息传递机制高性能计算知识虽然可能在实际中用得不多，但却可以使我们了解集群的基本工作机制。

云计算与大数据技术涉及面很广，本书在编写过程中参考并引用了大量前辈学者的研究成果和论述，对此编者向这些学者表示敬意，没有这些学者的努力本书是不可能完成的。云计算与大数据技术是一门高速发展的技术领域，新技术、新方法、新架构层出不穷，也是在不断探索和研究的新学科，由于作者的经验和能力所限，本书的结构、内容肯定存在许多疏漏和错误，望读者指正。

任课老师可以登录人民邮电出版社教育服务与资源下载社区（www.ryjiaoyu.com）下载本书的 PPT 课件、教学大纲、实验操作视频等教学资源，读者也可以登录本书支持网站 http://www.qhoa.org，获取相关支持。

编 者

2013 年 12 月

目 录 CONTENTS

第 7 章　分布式数据库——HBase　114

第 8 章　数据仓库平台——Hive　129

第 9 章　基于拓扑的流数据实时计算系统——Storm　149

第 10 章　云存储系统——Swift　165

PART 1

第 1 章 云计算基础

自从 2006 年谷歌公司 CEO 埃里克·施密特提出云计算概念后，云计算已经成为了全球关注度最高的 IT 词汇。随着信息技术水平的不断发展，云计算将会成为引领未来整个信息系统建设的主导者。

据 Occams Business 研究与咨询公司的预测，到 2020 年，全球云计算市场容量将从 2013 年的 900 亿美元增长到 6500 亿美元，复合增长率将达 29%。在云服务中，全球的平台即服务市场增长率最高，预计到 2020 年复合增长率将达 39%。在地理位置上，亚太地区是增长最快的地区，复合增长率将为 35%。云计算具有一体化的信息平台和运营平台，这种全新交付模式将会对 IT 界产生重大的影响，尤其是对那些传统的 IT 产业部门来说，该影响将是颠覆式的，其情形无异于给传统 IT 产业界带来了一场“地震”级的震撼。

1.1 云计算技术概述

1.1.1 云计算简介

云计算技术是硬件技术和网络技术发展到一定阶段而出现的一种新的技术模型，通常技术人员在绘制系统结构图时用一朵云的符号来表示网络，云计算因此而得名。云计算并不是对某一项独立技术的称呼，而是对实现云计算模式所需要的所有技术的总称。云计算技术的内容很多，包括分布式计算技术、虚拟化技术、网络技术、服务器技术、数据中心技术、云计算平台技术、分布式存储技术等；目前新出现的一些技术有 Hadoop、HPCC、Storm、Spark 等。从广义上说，云计算技术包括了当前信息技术中的绝大部分。

维基百科中对云计算的定义是：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求提供给计算机和其他设备，它就像我们日常生活中用水和用电一样，按需付费，而无需关心水、电是从何而来的。

2012 年的国务院政府工作报告将云计算作为国家战略性新兴产业给出了定义：云计算是基于互联网的服务的增加、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融合的产物，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。

对于以上的定义，我们可以从非技术的角度将云计算理解为它是一种通过网络的资源整合输出模式，只要是为了达到资源整合输出这个目的的技术都可以被称为云计算技术。从定义中也可以看出网络在云计算技术中的重要性，如果没有网络的高速发展，则云计算这种模式是无法实现的。

云计算技术的出现改变了信息产业传统的格局。传统的信息产业企业既是资源的整合者又是资源的使用者，这就像一个电视机企业既要生产电视机还要生产发电机一样，这种格局并不符合现代产业分工高度专业化的需求，同时也不符合企业需要灵敏地适应客户的需要。传统的计算资源和存储资源大小通常是相对固定的，不能及时响应客户需求的不断变化，企业的计算和存储资源要么是被浪费，要么是面对客户峰值需求时力不从心。

云计算时代的 3 种基本角色为资源的整合运营者、资源的使用者、终端客户。资源的整合运营者就像是发电厂一样负责资源的整合输出，资源的使用者负责将资源转变为满足客户需求的各种应用，终端客户为资源的最终消费者。

云计算技术使资源与用户需求之间是一种弹性化的关系，资源的使用者和资源的整合者并不是一个企业，资源的使用者只需要对资源按需付费，从而敏捷地响应客户不断变化的资源需求，这一方法降低了资源使用者的成本，提高了资源的利用效率。

云计算这种新的模式的出现被认为是信息产业的一大变革，从而吸引了大量企业的注意力。国际巨头 IBM、微软、谷歌、DELL 等企业都在云计算领域进行了全面的布局，变革之时正是机会出现的时候，云计算的出现更是给国内企业一次重新布局的机会，可以看到国内的华为、中兴、腾讯、阿里、联想、浪潮、五舟等企业都相继提出自己的云计算战略规划，并在云计算技术和市场都进行了全面的布局。

云计算技术作为一项涵盖面广且对产业影响深远的技术，未来将逐步渗透到信息产业和其他产业的方方面面，并将深刻改变产业的结构模式、技术模式和产品销售模式，进而深刻影响人们的生活。云计算会逐步成为人们生活中必不可少的技术。同时移动互联网的出现使云计算应用走向了人们的指间，推动了云计算技术的应用发展，今后云计算将是一项随时、随地、随身为我们提供服务的技术。云计算的出现也将如电的发现一般，为信息产业的发展提供无限的想象空间，使应用的创新能力得到完全释放。

1.1.2 云计算的特点

为了理解云计算这个概念，只了解一个简单的定义是不够的，我们还需要利用云计算技术的特点来判断一个技术是否是云计算技术。与传统的资源提供方向相比，云计算具有以下特点。

1．资源池弹性可扩张

云计算系统的一个重要特征就是资源的集中管理和输出，这就是所谓的资源池。从资源低效率的分散使用到资源高效的集约化使用正是云计算的基本特征之一。分散的资源使用方法造成了资源的极大浪费，现在每个人都可能有一到两台自己的计算机，但对这种资源的利用率却非常的低，计算机在大量时间都是在等待状态或是在处理文字数据等低负荷的任务。资源集中起来后资源的利用效率会大大地提高，随着资源需求的不断提高，资源池的弹性化扩张能力成为云计算系统的一个基本要求，云计算系统只有具备了资源的弹性化扩张能力才能有效地应对不断增长的资源需求。大多数云计算系统都能较为方便地实现新资源的加入。

2．按需提供资源服务

云计算系统带给客户最重要的好处就是敏捷地适应用户对资源不断变化的需求，云计算系统实现按需向用户提供资源能大大节省用户的硬件资源开支，用户不用自己购买并维护大量固定的硬件资源，只需向自己实际消费的资源量来付费。按需提供资源服务使应用开发者在逻辑上可以认为资源池的大小是不受限制的，这就使应用软件的开发者拥有了更大的想象

空间和创新空间，更多的有趣应用将在云计算时代被创造出来，应用开发者的主要精力只需要集中在自己的应用上。

3．虚拟化

现有的云计算平台的重要特点是利用软件来实现硬件资源的虚拟化管理、调度及应用。通过虚拟平台，用户使用网络资源、计算资源、数据库资源、硬件资源、存储资源等，与在自己的本地计算机上使用的感觉是一样的，相当于是在操作自己的计算机，而在云计算中利用虚拟化技术可大大降低维护成本和提高资源的利用率。

4．网络化的资源接入

从最终用户的角度看，基于云计算系统的应用服务通常都是通过网络来提供的，应用开发者将云计算中心的计算、存储等资源封装为不同的应用后往往会通过网络提供给最终的用户。云计算技术必须实现资源的网络化接入才能有效地向应用开发者和最终用户提供资源服务。这就像有了发电厂必须还要有输电线才能将电传送给用户。所以网络技术的发展是推动云计算技术出现的首要动力。目前一些企业将网络化的软件和硬件都称为云计算，就是因为网络化的资源接入方式是从最终用户角度能看到的云计算的重要特征之一，这些产品的称呼不一定准确，但却是对云计算特征的反映。

5．高可靠性和安全性

用户数据存储在服务器端，而应用程序在服务器端运行，计算由服务器端来处理。所有的服务分布在不同的服务器上，如果什么地方（节点）出问题就在什么地方终止它，另外再启动一个程序或节点，即自动处理失败节点，从而保证了应用和计算的正常进行。

数据被复制到多个服务器节点上有多个副本（备份），存储在云里的数据即使遇到意外删除或硬件崩溃也不会受到影响。

1.1.3 云计算技术分类

目前已出现的云计算技术种类非常多，云计算的分类可以有多种角度：从技术路线角度可以分为资源整合型云计算和资源切分型云计算；从服务对象角度可以被分为公有云和私有云、混合云和社区云；按资源封装的层次可以分为基础设施即服务（Infrastructure as a Service，IaaS）、平台即服务（Platform as a Service，PaaS）和软件即服务（Software as a Service，SaaS）。

1．按技术路线分类

（1）资源整合型云计算

这种类型的云计算系统在技术实现方面大多体现为集群架构，通过将大量节点的计算资源和存储资源整合后输出。这类系统通常能实现跨节点弹性化的资源池构建，核心技术为分布式计算和存储技术。MPI、Hadoop、HPCC、Storm 等都可以被分类为资源整合型云计算系统。

（2）资源切分型云计算

这种类型最为典型的就是虚拟化系统。这类云计算系统通过系统虚拟化实现对单个服务器资源的弹性化切分，从而有效地利用服务器资源。其核心技术为虚拟化技术。这种技术的优点是用户的系统可以不做任何改变接入采用虚拟化技术的云系统，是目前应用较为广泛的技术，特别是在桌面云计算技术上应用得较为成功；缺点是跨节点的资源整合代价较大。KVM、VMware 都是这类技术的代表。

2．按服务对象分类

（1）公有云（Public Cloud）

公有云是指面向公众的云计算服务，由云服务提供商运营。其目的是为终端用户提供从应用程序、软件运行环境，到物理基础设施等各种各样的IT资源。它对云计算系统的稳定性、安全性和并发服务能力有更高的要求。

（2）私有云（Private Cloud）

私有云是指企业自建自用的云计算中心，且具备许多公有云环境的优点。主要服务于某一组织内部的云计算服务，其服务并不向公众开放，如企业、政府内部的云服务。

（3）混合云（Hybrid Cloud）

混合云是把公有云和私有云结合在一起的方式。在这个模式中，用户通常将非企业关键信息外包，并在公有云上处理，而掌握企业关键服务及数据的内容则放在私有云上处理。

（4）社区云（Community Cloud）

社区云是公有云范畴内的一个组成部分。它由众多利益相仿的组织掌控及使用，其目的是实现云计算的一些优势，例如特定安全要求、共同宗旨等。社区成员共同使用云数据及应用程序。

目前，公有云引领着云市场，占据着大量的市场份额。采用公有云的一个主要原因是“按需付费”的成本效益模型。另外，它还通过优化运营、支持和维护服务给云服务供应商带来了规模经济。私有云市场使用规模公次于公有云，主要是因为它在安全性方面做得更好。混合云模型目前市场中占有份额较少，但未来发展空间巨大。社区云由于共同承担费用的用户数远比公有云少，因此也更贵，但隐私度、安全性和政策遵从都比公有云要高。用户可以根据其需求，选择一种适合自己的云计算模式。

3．按资源封装的层次分类

（1）基础设施即服务

把单纯的计算和存储资源不经封装地直接通过网络以服务的形式提供给用户使用。客户可以使用“基础计算资源”，如处理能力、存储空间、网络组件或中间件，并掌控操作系统、存储空间、已部署的应用程序及网络组件（如防火墙、负载平衡器等），但不掌控云基础架构。这类云计算服务用户的自主性较大，就像是自来水厂或发电厂一样直接将水电送出去。

这种方式可以满足非IT企业对IT资源的需求，同时还不需要花费大量资金购置服务器和雇佣更多的IT人员，使他们可以将自己的主要精力放在自己的主业上。同时，这种云服务还使用自动化技术来根据用户的业务量自动分配合适的服务器数量，用户不必为自己业务的扩展或者收缩而考虑IT资源是否合适。同时用户不必担心IT设施的折旧问题，只需根据自己的服务器使用量交付月租金即可。这类云服务的对象往往是具有专业知识能力的资源使用者，传统数据中心的主机租用等可能作为IaaS的典型代表。

（2）平台即服务

计算和存储资源经封装后，以某种接口和协议的形式提供给用户调用，资源的使用者不再直接面对底层资源。即资源的使用者不需要管理或控制底层的云基础设施，包括网络、服务器、操作系统、存储等；但客户能控制部署的应用程序，也可能控制运行应用程序的托管环境配置。PaaS位于云计算的中间层，主要面向软件开发者或软件开发商，提供基于互联网的软件开发测试平台。软件开发人员可以通过基于Web等技术直接在云端编写自己的应用程序，同时也可以将自己的应用程序托管到这个平台上。例如，Google的App Engine就是一个

可伸缩的 Web 应用程序开发和托管平台，开发者可以在其平台上开发出自己的 Web 程序并发布，而不需要担心自己的服务器能否承担未知的访问量，这样的平台得到了一些小型创业企业的青睐。

另外，这样的云平台还提供大量的 API 或者中间件供程序开发者使用，大大缩短了程序开发的周期；同时，程序代码存储在云端可以很方便联合开发。最重要的是用户不必再担心自己发布的应用需要多少硬件支持，因为，云端可以满足一切。

（3）软件即服务

将计算和存储资源封装为用户可以直接使用的应用，并通过网络提供给用户。SaaS 面向的服务对象为最终用户，用户只是对软件功能进行使用，无需了解任何云计算系统的内部结构，也不需要用户具有专业的技术开发能力。软件即服务是一种服务观念的基础。软件服务供应商以租赁的概念提供客户服务，而非购买。比较常见的模式是提供一组账号密码。

SaaS 相对 IaaS、PaaS 来说应该不会太陌生，例如，和我们日常生活相关的微信、飞信、QQ 等都有对应 Web 版本，我们也不必担心软件的更新和维护等问题，只需通过 Web 就可以获得相应的服务。也许用户对于像 QQ 这类的小软件来说并不能完全体会到 SaaS 的优势，但对于那些中小型企业和他们需要的 ERP、CRM 等来说，SaaS 是一种福音。首先，企业不必花费巨额资金购买软件的使用权；其次，企业也不必花费资金构建机房和雇佣人员；再次，企业也不必考虑机器折旧和软件升级维护等问题。

如图 1-1 所示，云计算系统按资源封装的层次分为 IaaS、PaaS、SaaS，分为对底层硬件资源不同级别的封装，从而实现将资源转变为服务的目的。传统的信息系统资源的使用者通常是以直接占有物理硬件资源的形式来使用资源的；而云计算系统通过 IaaS、PaaS、SaaS 等不同层次的封装将物理硬件资源封装后，以服务的形式利用网络提供给资源的使用者。在这里，资源的使用者可能是资源的二次加工者，也可能是最终应用软件的使用者。通常 IaaS、PaaS 层面向的资源使用者往往是资源的二次加工者。这类资源的使用者并不是资源的最终消费者，他们将资源转变为应用服务程序后，以 SaaS 的形式提供给资源的最终消费者。实现对物理资源封装的技术并不是唯一的，目前不少的软件都能实现，甚至有的系统只有 SaaS 层，并没有进行逐层的封装。

云计算的服务层次是根据服务类型即服务集合来划分的，与大家熟悉的计算机网络体系结构中层次的划分不同。在计算机网络中每个层次都实现一定的功能，层与层之间有一定关联。而云计算体系结构中的层次是可以分割的，即某一层次可以单独完成一项用户的请求而不需要其他层次为其提供必要的服务和支持。

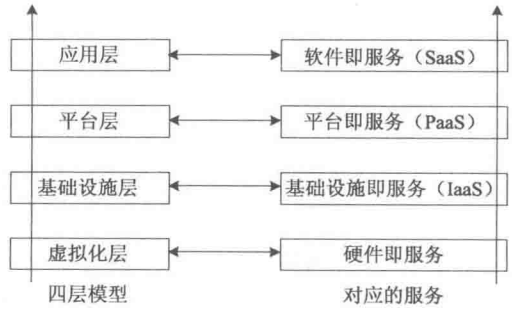

图 1-1 云计算服务体系结构

在云计算服务体系结构中各层次与相关云产品对应。

应用层对应 SaaS 软件即服务，如 Google APPS、SoftWare+Services、Microsoft CRM。

平台层对应 PaaS 平台即服务，如 IBM IT Factory、Google APP Engine、Force.com。

基础设施层对应 IaaS 基础设施即服务，如 Amazon EC2、IBM Blue Cloud、Rackspace。

虚拟化层对应硬件即服务结合 PaaS 提供硬件服务，包括服务器集群及硬件检测等服务。

1.1.4 计算机技术向现代信息技术演进的历程

回顾计算机技术的发展历程，可以清晰地看到计算机技术从面向计算逐步转变到面向数据的过程。从面向计算到面向数据是技术发展的必然趋势，并不能把云计算的出现归功于任何的个人和企业。这一过程的描述如图 1-2 所示，该图以时间为顺序对硬件、网络和云计算的演进过程进行了纵向和横向的对比。

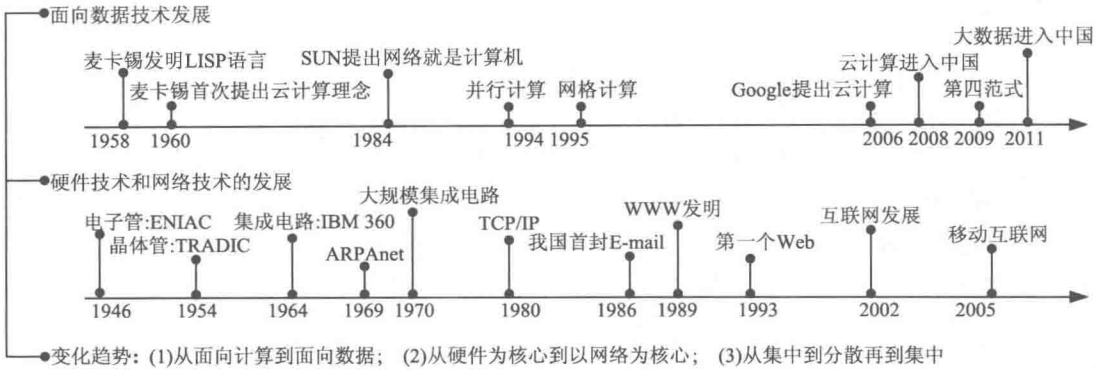

图 1-2 计算机技术向现代信息技术的演进

从图 1-2 中可以看到，在计算机技术的早期，由于硬件设备体积庞大，价格昂贵，数据的产生还是“个别”人的工作。这个时期的数据生产者主要是科学家或军事部门，他们更关注计算机的计算能力，计算能力的高低决定了研究能力和一个国家军事能力的高低。相对而言，由于这时数据量很小，数据在整个计算系统中的重要性并不突出。这时网络还没有出现，推动计算技术发展的主要动力是硬件的发展。这个时期是硬件的高速变革时期，硬件从电子管迅速发展到大规模集成电路。1969 年 ARPAnet 的出现改变了整个计算机技术的发展历史，网络逐步成为推动技术发展的一个重要力量。1989 年蒂姆·伯纳斯·李发明的万维网改变了信息的交流方式，特别是高速移动通信网络技术的发展和成熟，使现在数据的生产成为全球人的共同活动。人们生产数据不再是在固定时间和固定地点进行，而是随时随地都在产生数据。微博、博客、社交网、视频共享网站、即时通信等媒介随时都在生产着数据并被融入全球网络中。

从云计算之父约翰·麦卡锡提出云计算的概念，到大数据之父詹姆斯·尼古拉·格雷等人提出科学研究的第四范式，时间已经跨越了半个世纪。以硬件为核心的时代也是面向计算的时代。那时数据的构成非常简单，数据之间基本没有关联性，物理学家只处理物理实验数据，生物学家只处理生物学数据，计算和数据之间的对应关系是非常简单和直接的。这个时期研究计算和存储的协作机制并没有太大的实用价值。到了以网络为核心的时代数据的构成变得非常复杂，数据来源多样化，不同数据之间存在大量的隐含关联性。这时计算所面对的数据变得非常复杂，如社会感知、微关系等应用将数据和复杂的人类社会运行相关联，由于人人都是数据的生产者，人们之间的社会关系和结构就被隐含到了所产生的数据之中。数据的产生目前呈现出了大众化、自动化、连续化、复杂化的趋势。云计算、大数据概念正是在这样的一个背景下出现的。这一时期的典型特征就是计算必须面向数据，数据是架构整个系统的核心要素，这就使计算和存储的协作机制研究成为需要重点关注的核心技术，计算能有效找到自己需要处理的数据，可以使系统能更高效地完成海量数据的处理和分析。云计算和大数据这两个名词也可看作是描述了面向计算时代信息技术的两个方面，云计算侧重于描述