



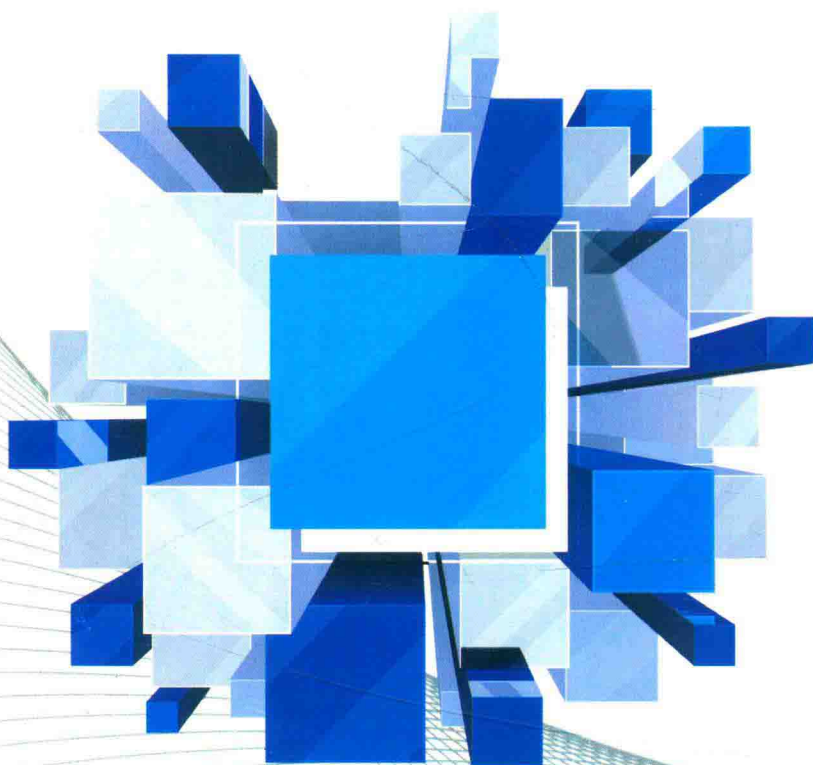
会物联网专家委员会推荐

教育物联网工程专业“十三五”规划教材

云计算及其实践教程 (第二版)

A Course of Cloud Computing and Its Practice (The Second Edition)

郝卫东 王志良 编著
刘宏岚 王 宁



西安电子科技大学出版社
<http://www.xduph.com>

课件 & 素材

物联网专家委员会推荐

普通高等教育物联网工程专业“十三五”规划教材

云计算及其实践教学

(第二版)

郝卫东 王志良 刘宏岚 王 宁 编著



西安电子科技大学出版社

内 容 简 介

本书系统、全面地介绍了面向教学的云计算理论、平台和应用实践。云计算理论部分主要包括 SPI 服务模型、虚拟化、云存储、云安全、云标准化、云计算与物联网通信等方面的知识；平台部分主要包括 Google、Amazon、OpenStack、Windows Azure、Hadoop、Spark 等主流云平台；应用实践部分主要包括基于 Windows Azure 和 Hadoop、Spark 的九个实验，并给出了云实践的路径建议。附录包括习题答案、习题增补和中英文术语对照表等内容。全书既重视基本概念、基本理论的阐述，也重视主流产品、应用方案、编程实现的介绍。在出版社网站上提供本书教学和实验所用的课件与相关素材。

本书可作为高等学校物联网工程、通信与信息系统、电子科学与技术、电子工程、计算机等专业的本科生教材，也可作为相近专业的教学参考书。

图书在版编目(CIP)数据

云计算及其实践教程/郝卫东等编著. —西安: 西安电子科技大学出版社, 2017.9

(普通高等教育物联网工程专业“十三五”规划教材)

ISBN 978-7-5606-4512-4

I. ① 云… II. ① 郝… III. ① 云计算—教材 IV. ① TP393.027

中国版本图书馆 CIP 数据核字(2017)第 144369 号

策 划 毛红兵

责任编辑 王 斌 毛红兵

出版发行 西安电子科技大学出版社(西安市太白南路2号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfb001@163.com

经 销 新华书店

印刷单位 陕西天意印务有限责任公司

版 次 2017年9月第2版 2017年9月第2次印刷

开 本 787毫米×1092毫米 1/16 印 张 28.5

字 数 675千字

印 数 3000~6000册

定 价 59.00元

ISBN 978-7-5606-4512-4/TP

XDUP 4804002-2

如有印装问题可调换

本社图书封面为激光防伪覆膜，谨防盗版。

前 言

由于云计算的发展十分迅猛，自《云计算及其实践教程》出版以来，新的 Hadoop 2.0 平台及 YARN 框架的出现、新的云平台 Spark 的出现及迅速被广泛接受，促使我们下决心更新教材内容，赶上时代发展。

根据教学反馈，学生对云计算实践很关注。原有教材主要关注在 Windows 平台的云开发，少量涉及 Linux 平台，修订版增加了 Linux 平台的云开发，并在宏观意义上给出了几条不同的加强云平台实践和开发的学习路径。

本书共分 11 章。第 1 章云计算概论，介绍了云计算的定义和特征、云计算的 SPI 服务模式；第 2 章主流云平台，介绍了 Amazon 云平台、Google 云平台、OpenStack 云平台等；第 3 章 Windows Azure 云平台，介绍了微软的云服务、Windows Azure 云平台及其组成；第 4 章虚拟化，介绍了虚拟化的概念、功能、产品和应用方案；第 5 章 Hadoop 云平台，介绍了并行计算的概念、Hadoop 的各个组件构成和原理、Hadoop 程序实例运行；第 6 章 Spark 平台，介绍 Spark 产生背景、生态系统、核心概念 RDD、程序设计实例；第 7 章云存储，介绍了云存储的概念、FCoE 接口、NoSQL 的概念和实例、云存储方案设计和存储虚拟化；第 8 章云安全，介绍了云安全的概念、技术和应用方案；第 9 章云标准，介绍了云计算的各种标准化组织；第 10 章云计算与物联网通信，介绍了物联网三层体系结构、物联网通信的概念、WPAN、WLAN、WMAN 和 WWAN 等；第 11 章云计算实践，介绍了基于 Windows Azure 和 Hadoop、Spark 平台的九个实验的实验指导，其实验步骤详细，初学者也可以参照学习和实践。本书包括三个附录：附录 1 为各章习题答案；附录 2 为增补习题及其答案；附录 3 为中英文术语对照表。本书提供教学用的 PPT 课件、实验源程序和实验用的素材等，可在出版社网站下载。

与第一版相比，第二版修订的内容包括：

第一，新增了云计算平台 Spark 的基本理论，包括产生背景、生态系统、核心概念 RDD、程序设计实例。新增 YARN 的核心概念以及架构。

第二，补充完善了虚拟机的迁移、网络虚拟化、桌面虚拟化等方面的内容。

第三，NoSQL 技术目前应用越来越广泛，包括阿里云等国内厂商也在迅速跟进使用 MongoDB 等 NoSQL 数据库，学生毕业设计和就业已开始涉及这些新知识。更新原有教材，加强 NoSQL 的实例和分类以及 FCoE 接口、集群存储的概念和方案。

第四，在实践方面，增加了 Linux 平台的云开发，包括支持 YARN 的 Hadoop 分布式安装和部署、Spark 安装、编写和运行 Hadoop 和 Spark 程序、Spark Shell 程序命令执行等，并给出加强云平台实践和开发的学习路径建议。

第五，各章中更新和增加了“温馨提示”部分，帮助读者理解易混淆的相关概念。

第六，增加了一些计算题和设计题，并集中放置到附录 2 中，这些题目关注并行计算和分布式计算、资源管理和调度等基本理论问题，有助于学生深入思考云计算背后的科学问题。

与现有的云计算图书比较，本书的特点在于：

第一，面向物联网加以阐述，切合专业。云计算是物联网的重要基础，物联网是云计算的应用之一。本书在内容上包括云计算与物联网的关系、云计算与物联网通信、可以用于物联网领域中的数据挖掘和数据分析的 Mahout 开源平台、Spark MLlib 开源平台等内容。

第二，注重基础知识和基本理论，形成体系，便于教学。本书每个章节都有“温馨提示”部分，以使读者理解易混淆的概念。

第三，从实践入手，使读者有获得感。例如，听视频、看电子书、自学课件中简单的非关键部分；在网上申请阿里云、百度云、Amazon 云等账号(这三个都易于申请且有免费的时间段赠送)；自己动手安装 Hadoop、Spark 运行环境并用 Java 或 Scala 语言开发 Linux 平台下的云程序；自己动手安装 Windows Azure 运行环境并用 C#语言开发 Windows 平台下的云程序；自己动手安装 VMware Workstation 虚拟化软件或 OpenStack 开源虚拟化管理平台，创建自己的虚拟机。

第四，关注云计算领域的热点。本书不仅关注 Mahout、Spark、MLlib 等大数据和机器学习方面的内容，也关注网络虚拟化、桌面虚拟化等云网络方面的内容，并创新性地以习题形式关注并行计算和分布式计算、资源管理和资源调度等基本理论问题。

本书的初稿在北京科技大学计算机与通信工程学院物联网专业进行了多次试讲，并指导学生做了书中的全部实验，收到了良好的效果。本书的出版得到了教育部“本科教学工程”、“专业综合改革试点”项目经费和北京科技大学教材建设基金的资助，在此表示感谢。

本书可作为高等学校物联网工程、通信与信息系统、电子科学与技术、电子工程、计算机等专业的本科生教材，也可作为相近专业的教学参考书。

郝卫东编写了本书第 3、4、5、6、11 章，王志良编写了第 1、10 章，刘宏岚编写了第 7、8、9 章和附录 1、2、3，王宁编写了第 2 章。

感谢刘宏岚整理了全书所有的图片。感谢王志良教授对本书的指导意见，让我们重视习题和实验，并联系了微软公司对我们进行指导。感谢西安电子科技大学出版社毛红兵编辑和王斌编辑的大力支持和认真细致的工作。本书部分引用了相关参考文献和网络资源，在此，我们对这些资料的作者们表示衷心感谢。

由于编者水平有限，书中难免存在不妥之处，恳请广大读者批评指正。

作者联系方式：郝卫东，E-mail: Wed@ustb.edu.cn。

编 者

2017 年 5 月

第一版前言

随着云计算的深入发展,高校的物联网工程专业急需一本云计算及其实践方面的教材。本书的编写目的就是为了比较系统、全面地介绍面向物联网的云计算理论和技术基础,相应的主流应用平台,产品的体系结构、设计方法、实际运行方案等方面的知识,以及该领域国际上比较新的热点课题,使相关专业的学生及从事此方向研究的工程技术和科研人员,在理论和实际工程应用方面得到较为有价值的训练。

本书共分 10 章。第 1 章云计算概论,介绍了云计算的定义和特征、云计算的 SPI 服务模型;第 2 章主流云平台,介绍了 Amazon 云平台、Google 云平台、OpenStack 云平台等;第 3 章 Windows Azure 云平台,介绍了微软的云服务、Windows Azure 云平台及其组成;第 4 章虚拟化,介绍了虚拟化的概念、产品和应用方案;第 5 章基于 Hadoop 的云编程,介绍了并行计算的概念、Hadoop 的各个组件构成和原理、Hadoop 的程序实例运行;第 6 章云存储,介绍了云存储的概念、NoSQL 的概念和实例、方案设计和存储虚拟化;第 7 章云安全,介绍了云安全的概念、技术和应用方案;第 8 章云标准,介绍了云计算的各种标准化组织;第 9 章云计算与物联网通信,介绍了物联网三层体系结构、物联网通信的概念、WPAN、WLAN、WMAN 和 WWAN 等;第 10 章 Windows Azure 和 Hadoop 实验,介绍了基于 Windows Azure 和 Hadoop 平台的六个实验的实验指导,其实验步骤详细,初学者也可以参照学习和实践。本书包括两个附录:附录 1 为习题答案;附录 2 为中英文术语对照表。本书还附有光盘,提供教学用的 PPT 课件、实验源程序和实验用的素材等。

与现有的云计算图书比较,本书的首要特色及创新之处在于面向物联网加以阐述,重视在理论分析基础上的应用开发和编程实验,重点包括基于 Windows Azure 和 Hadoop 云平台的开发实验的详细过程。本书的第二个特色是追踪前沿热点问题,如基于 Hadoop 的云编程、OpenStack 云平台等。其中基于 Hadoop 的 Mahout 开源平台可用于物联网领域中的数据挖掘和数据分析。本书的第三个特色是面向基础。本书从基本的知识讲起,帮助学生有效地学习和掌握基础知识。书中每个章节都有“温馨提示”部分,对相关知识的背景和基础加以介绍。本书设置了形式多样的习题,以帮助学生巩固所学的知识。

本书的初稿在北京科技大学计算机与通信工程学院物联网专业进行了两次试讲,并指导学生做了书中全部的实验,收到了良好的效果。北京科技大学是教育部批准的首批物联网本科专业的 28 所学校之一(2010 年),是教育部批准的首批物联网本科特色专业的 6 所学校之一(2011 年)。北京科技大学物联网专业与微软公司、IBM 公司等有着长期教学研究合作计划,同时得到国家级高等学校(物联网工程)特色专业建设重点项目、北京市支持中央在京高校共建项目、北京科技大学教改重点项目的支持和资助,在此一并表示感谢。本书的出版得到了教育部“本科教学工程”“专业综合改革试点”项目经费和北京科技大学教材建设基金的资助,在此也表示感谢。

本书可以作为高等学校物联网工程、通信与信息系统、电子科学与技术、电子工程、计算机等专业的本科生教材,也可以作为相近专业的教学参考书。

郝卫东编写了本书第 3、4、5、10 章，王志良编写了第 1、9 章，刘宏岚编写了第 6、7、8 章和附录 1、2，王宁编写了第 2 章。王丹丹、陈宇、马欢、孔妍、习荣华、耿绪超、迪娜·艾德力汗、于汝云和覃小娜参与了部分章节的编写。

感谢刘宏岚整理了全书所有的图片。感谢王志良教授对本书的意见，让我们重视习题和实验，并联系了微软公司对我们进行指导。感谢西安电子科技大学出版社毛红兵编辑和王斌编辑的大力支持和认真细致的工作。本书部分引用了相关参考文献和网络资源，在此，我们对这些资料的作者们表示衷心感谢。

由于编者水平有限，书中难免存在不妥之处，恳请广大读者批评指正。

作者联系方式：郝卫东，E-mail: Wed@ustb.edu.cn。

编 者

2014 年 3 月

目 录

第 1 章 云计算概论	
1.1 云计算的定义和特征	1
1.1.1 云计算的定义	1
1.1.2 云计算的特征	3
1.1.3 云计算系统的组成	3
1.1.4 云计算的部署模式	4
1.2 云计算应用实例	5
1.3 服务理论	9
1.3.1 服务的概念	9
1.3.2 几种常见的服务	10
1.3.3 服务概念的特征	11
1.3.4 面向服务的开发	12
1.4 云计算的 SPI 服务模型	14
1.4.1 IaaS	14
1.4.2 PaaS	15
1.4.3 SaaS	17
1.4.4 IaaS、PaaS 和 SaaS 的比较	18
1.5 云计算与相关领域的关系	18
1.5.1 云计算与网格计算的关系	18
1.5.2 云计算与 P2P 计算的关系	20
1.5.3 云计算与集群计算的关系	22
1.5.4 云计算与物联网的关系	23
本章小结	24
习题与思考	24
第 2 章 主流云平台	26
2.1 云平台综述	26
2.2 Amazon 云平台	26
2.2.1 弹性计算云 EC2	27
2.2.2 简单存储服务 S3	33
2.2.3 简单数据库服务 Simple DB	35
2.2.4 简单队列服务 SQS	37
2.3 Google 云平台	39
2.3.1 Google 文件系统 GFS	40
2.3.2 分布式计算编程模型 MapReduce	42
2.3.3 分布式锁服务 Chubby	43
2.3.4 分布式数据存储系统 Bigtable	45
2.3.5 Google App Engine	49
2.4 OpenStack 云平台	51
2.4.1 计算服务 Nova	54
2.4.2 存储服务 Swift	58
2.4.3 镜像服务 Glance	63
2.4.4 身份服务 Keystone	63
2.4.5 用户界面服务 Horizon	64
本章小结	65
习题与思考	65
第 3 章 Windows Azure 云平台	67
3.1 微软云计算服务概述	67
3.1.1 面向消费者的云服务	68
3.1.2 面向企业的云服务	69
3.1.3 平台发展目标	70
3.2 Windows Azure 平台简介	70
3.3 云操作系统 Windows Azure	71
3.3.1 Windows Azure 的组成	71
3.3.2 Windows Azure 计算服务	72
3.3.3 Windows Azure 存储服务	74
3.3.4 Windows Azure Fabric 控制器	78
3.3.5 Windows Azure 应用场景	80
3.4 SQL Azure	81
3.4.1 SQL Azure 概述	81
3.4.2 SQL Azure 数据库体系结构	82
3.4.3 SQL Azure 数据库和 SQL Server 数据库的对比	84
3.5 Windows Azure AppFabric	86
3.5.1 Windows Azure AppFabric 概述	86
3.5.2 服务总线	87
3.5.3 访问控制服务	90
3.5.4 分布式缓存	90
本章小结	91
习题与思考	92

第4章 虚拟化	93	5.1.2 并行计算的体系结构	128
4.1 虚拟化概述	93	5.1.3 集群计算	130
4.2 服务器虚拟化	95	5.1.4 并行计算的进程模型	133
4.2.1 服务器虚拟化概述	95	5.1.5 并行编程模型	134
4.2.2 服务器虚拟化的类型	95	5.2 Hadoop 概述	136
4.2.3 服务器虚拟化的架构	99	5.2.1 Hadoop 的由来	136
4.2.4 服务器虚拟化的核心技术	99	5.2.2 Hadoop 的特点	136
4.3 虚拟化的主要功能	100	5.2.3 Hadoop 的基本结构	137
4.3.1 虚拟机的基本功能	100	5.2.4 Hadoop 的应用	138
4.3.2 虚拟机的迁移	102	5.3 HDFS	138
4.3.3 虚拟化应用举例	103	5.3.1 HDFS 的功能	138
4.4 服务器虚拟化主流厂商及产品	105	5.3.2 HDFS 的结构	138
4.4.1 VMware ESX	105	5.3.3 HDFS 文件读/写操作流程	139
4.4.2 Citrix XenServer	106	5.3.4 HDFS 如何实现可靠存储、 副本管理	141
4.4.3 Microsoft Hyper-V	108	5.4 MapReduce	142
4.4.4 RedHat KVM	109	5.4.1 MapReduce 原理	142
4.4.5 主流虚拟化产品的比较	110	5.4.2 MapReduce 执行流程	143
4.5 服务器虚拟化应用方案设计	112	5.4.3 MapReduce 数据流程	144
4.5.1 需求分析	112	5.4.4 MapReduce 的容错机制	145
4.5.2 方案准备	113	5.5 YARN	146
4.5.3 方案设计	114	5.5.1 YARN 是一个资源管理平台	146
4.5.4 方案实施	115	5.5.2 原 MapReduce 框架存在的问题	147
4.5.5 方案效益	116	5.5.3 YARN 架构	148
4.6 网络虚拟化	117	5.5.4 YARN 工作流程	148
4.6.1 传统的网络虚拟化	117	5.5.5 YARN 框架相对于旧的 MapReduce 框架的优势	149
4.6.2 虚拟以太网交换机 VEB	118	5.6 HBase	149
4.6.3 VEPA 和 VN-Tag 技术	119	5.6.1 HBase 概述	149
4.7 桌面虚拟化	122	5.6.2 HBase 与关系型数据库的比较	150
4.7.1 桌面虚拟化的概念和技术	122	5.6.3 HBase 的数据模型	150
4.7.2 网络显示协议及其实例	123	5.6.4 HBase Shell 命令的应用	150
4.7.3 桌面虚拟化实例	124	5.7 Zookeeper	152
4.8 应用虚拟化	125	5.7.1 Zookeeper 的功能	152
4.8.1 应用虚拟化概述	125	5.7.2 Zookeeper 的数据模型	152
4.8.2 应用虚拟化实例	125	5.7.3 Zookeeper 的典型应用场景	153
本章小结	126	5.8 Hadoop 的程序实例运行与分析	155
习题与思考	126	5.8.1 WordCount 实例	155
第5章 Hadoop 云平台	128	5.8.2 每年最高气温实例	158
5.1 并行计算	128		
5.1.1 并行计算概述	128		

5.8.3 基于 Hadoop 的数据挖掘
 开源平台——Mahout..... 160
本章小结..... 162
习题与思考..... 162
第 6 章 Spark 平台..... 164
6.1 三种计算框架..... 164
 6.1.1 批处理(Batch)计算..... 165
 6.1.2 流式(Streaming)计算..... 165
 6.1.3 交互式(Interactive)计算..... 165
6.2 Spark 产生背景..... 166
6.3 Spark 特点..... 166
 6.3.1 高效..... 167
 6.3.2 易用..... 168
 6.3.3 与 Hadoop 集成..... 169
6.4 Spark 生态系统..... 170
 6.4.1 Spark 生态系统概述..... 170
 6.4.2 Alluxio..... 171
 6.4.3 Mesos 和 YARN..... 172
 6.4.4 Shark 和 Spark SQL..... 172
 6.4.5 Spark Streaming..... 172
 6.4.6 GraphX..... 173
 6.4.7 MLBase 和 MLlib..... 173
6.5 Spark 核心概念 RDD..... 174
 6.5.1 Spark 的核心概念..... 174
 6.5.2 利用本地文件或 HDFS 文件
 创建 RDD..... 174
 6.5.3 对 RDD 进行操作..... 174
 6.5.4 RDD Transformation 举例..... 176
 6.5.5 RDD Action 举例..... 177
 6.5.6 Key/Value 类型的 RDD..... 177
6.6 Spark 程序设计实例..... 178
 6.6.1 实例 1: WordCount..... 178
 6.6.2 Spark 程序设计的基本流程..... 180
 6.6.3 Spark 程序设计的 Scala 语言..... 180
 6.6.4 实例 2: 用蒙特卡洛算法
 分布式估算 Pi..... 180
 6.6.5 程序架构及相关概念..... 182
 6.6.6 体验 Spark 交互式模式
 Spark-shell..... 183

6.6.7 提交 Spark 程序..... 183
6.7 进一步理解 Spark 核心概念 RDD..... 185
 6.7.1 RDD 与 DAG..... 185
 6.7.2 划分 Stage..... 185
 6.7.3 划分 Stage 举例..... 186
6.8 进一步理解 Spark 新概念..... 187
 6.8.1 Dataset 的概念和使用..... 188
 6.8.2 SparkSession 的概念和使用..... 189
本章小结..... 191
习题与思考..... 191
第 7 章 云存储..... 193
7.1 云存储概述..... 193
 7.1.1 云存储的概念..... 193
 7.1.2 云存储的结构模型..... 194
 7.1.3 云存储国内外发展现状..... 195
 7.1.4 云存储相比传统存储的优势..... 197
7.2 存储结构..... 198
 7.2.1 DAS(直接连接存储)..... 198
 7.2.2 NAS(网络附加存储)..... 198
 7.2.3 SAN(存储区域网络)..... 199
 7.2.4 集群存储..... 200
7.3 存储设备..... 203
 7.3.1 存储设备概述..... 203
 7.3.2 磁盘阵列(RAID)..... 203
7.4 存储接口..... 206
 7.4.1 SCSI 接口..... 206
 7.4.2 FC 接口..... 208
 7.4.3 iSCSI 接口..... 211
 7.4.4 InfiniBand 接口..... 212
 7.4.5 Myrinet 接口..... 213
 7.4.6 FCoE 接口..... 214
7.5 NoSQL 数据库..... 216
 7.5.1 数据库的分类和 NoSQL 简介..... 216
 7.5.2 关系数据库的问题和 NoSQL 的
 出现..... 218
 7.5.3 NoSQL 的特点..... 219
 7.5.4 NoSQL 的实例..... 219
 7.5.5 NoSQL 的常见数据结构..... 220
7.6 云存储上传和下载文件的设计..... 221

7.6.1 概要设计	221	9.3.14 IEEE.....	249
7.6.2 MySQL 数据库设计	222	9.3.15 ITU-T.....	249
7.6.3 详细设计	223	本章小结	250
7.7 存储虚拟化	225	习题与思考	250
7.7.1 存储虚拟化的概念与分类	225	第 10 章 云计算与物联网通信	251
7.7.2 服务器级别的存储虚拟化	226	10.1 物联网三层体系结构	251
7.7.3 存储设备级别的存储虚拟化	226	10.1.1 感知层关键技术	252
7.7.4 存储网络级别的存储虚拟化	227	10.1.2 网络层关键技术	254
本章小结	228	10.1.3 应用层关键技术	254
习题与思考	228	10.2 物联网通信概述	255
第 8 章 云安全	230	10.3 ZigBee 技术	258
8.1 云安全概述	230	10.3.1 ZigBee 技术的来源与优势	258
8.1.1 云安全的定义	230	10.3.2 ZigBee 技术的协议架构	259
8.1.2 云安全与传统网络安全的差别	231	10.3.3 ZigBee 技术在物联网中的应用	262
8.1.3 云安全发展现状	231	10.4 蓝牙(Bluetooth)技术	265
8.2 云安全技术	232	10.4.1 蓝牙技术的来源与特点	265
8.2.1 灾难备份和恢复	232	10.4.2 蓝牙技术的应用及产品	269
8.2.2 可信计算	235	10.5 超宽带(UWB)技术	270
8.2.3 云支付	237	10.5.1 超宽带的定义	270
8.2.4 应用方案和设计实例	240	10.5.2 超宽带技术的特点与应用	272
本章小结	242	10.5.3 超宽带技术的两大技术标准	274
习题与思考	242	10.5.4 超宽带技术与其他无线 通信技术的比较	276
第 9 章 云标准	243	10.6 60 GHz 通信技术	277
9.1 云计算标准化的意义	243	10.6.1 60 GHz 通信技术的特点	277
9.2 云计算标准化的现状	244	10.6.2 60 GHz 标准化进程	279
9.3 云计算标准化组织	244	10.6.3 60 GHz 组网中的非视距传输	281
9.3.1 美国国家标准与技术研究院	244	10.7 无线 LAN 通信技术	282
9.3.2 开放云计算联盟	245	10.7.1 无线 LAN 通信技术的标准	282
9.3.3 分布式管理任务组	245	10.7.2 无线 LAN 通信技术的 应用和组网	285
9.3.4 企业云买方理事会	246	10.8 无线 MAN 通信技术	287
9.3.5 云安全联盟	246	10.8.1 WiMAX 的概念和特点	287
9.3.6 《云开放宣言》	246	10.8.2 WiMAX 的演进	288
9.3.7 存储网络工业协会	247	10.8.3 WiMAX 系统的结构	289
9.3.8 欧洲电信标准协会	247	10.9 移动通信网	289
9.3.9 开放网格论坛	247	10.9.1 移动通信网的基本组成	289
9.3.10 开放云计算工作组	248	10.9.2 移动通信网络的发展历程	290
9.3.11 云计算互操作论坛	248	10.9.3 WCDMA 技术	294
9.3.12 电信管理论坛	248		
9.3.13 ISO/IEC	249		

10.9.4	CDMA 2000 技术.....	295
10.9.5	TD-SCDMA 技术.....	296
10.9.6	LTE 技术.....	297
本章小结		299
习题与思考		299
第 11 章 云计算实践 303		
11.1	建立和启动 Windows Azure 程序	
	开发环境	303
11.1.1	实验目的.....	303
11.1.2	实验环境.....	303
11.1.3	实验内容.....	303
11.1.4	上机思考题.....	306
11.2	创建 Windows Azure Web 角色	
	应用程序	306
11.2.1	实验目的.....	306
11.2.2	实验环境.....	307
11.2.3	实验内容.....	307
11.2.4	上机思考题.....	315
11.3	编写 WCF 云后台辅助角色	
	应用程序	315
11.3.1	实验目的.....	315
11.3.2	实验环境.....	315
11.3.3	实验原理.....	315
11.3.4	实验内容.....	316
11.3.5	上机思考题.....	327
11.4	编写 Table 存储服务应用程序	328
11.4.1	实验目的.....	328
11.4.2	实验环境.....	328
11.4.3	实验原理.....	328
11.4.4	实验内容.....	329
11.4.5	上机思考题.....	347
11.5	编写基于 Blob 的云存储应用程序	348
11.5.1	实验目的.....	348
11.5.2	实验环境.....	348
11.5.3	实验内容.....	348
11.5.4	上机思考题.....	358
11.6	Hadoop 的伪分布式部署.....	358
11.6.1	实验目的.....	358
11.6.2	实验环境.....	358
11.6.3	实验内容.....	359
11.6.4	上机思考题.....	366
11.7	支持 YARN 的 Hadoop 在两个虚拟机中	
	分布式运行	366
11.7.1	实验目的.....	366
11.7.2	实验环境	366
11.7.3	实验原理.....	366
11.7.4	实验内容.....	367
11.7.5	上机思考题.....	383
11.8	Spark 安装部署及上机操作	383
11.8.1	实验目的.....	383
11.8.2	实验环境.....	383
11.8.3	实验内容.....	384
11.8.4	上机思考题.....	388
11.9	云中的 Spark 实验	388
11.9.1	实验目的.....	388
11.9.2	实验环境.....	389
11.9.3	实验内容.....	389
11.9.4	上机思考题.....	394
11.10	云实践路径推荐	394
11.10.1	从一份调查问卷谈云实	
	践路径	394
11.10.2	结合翻转课堂进行云实践.....	396
11.10.3	亚马逊 AWS 云服务的	
	申请步骤	397
11.10.4	在 VMWare 上安装 Linux	
	虚拟机	403
附录 1 习题答案		407
附录 2 增补习题及其答案		415
附录 3 中英文术语对照表		432
参考文献		441



第1章 云计算概论

本章要点:

- 云计算的定义
- 云计算的特征
- 云计算系统的组成
- 服务理论
- 云计算的 SPI 服务模型
- 云计算与相关领域的关系



课件

1.1 云计算的定义和特征

1.1.1 云计算的定义

虽然云计算技术发展迅速,应用日渐广泛,但是目前,仍没有一个关于云计算的统一定义。通常得到较多认可的云计算(Cloud Computing)的定义有:

云计算是一种 IT 服务的交付和使用模式,即用户通过网络以按需、易扩展的方式获得所需的资源(硬件、平台、软件),提供资源的网络被称为“云”。

☛ 温馨提示:

云计算定义中所说的“资源(硬件、平台、软件)”,分别对应 SPI 服务模型中的“基础设施作为服务(IaaS)”、“平台作为服务(PaaS)”、“软件作为服务(SaaS)”。SPI 服务模型见 1.4 节。

“云”中的资源在使用者看来是可以无限扩展的,并且可以随时获取,按需使用,按使用方式付费。“云”的这种特性使其成为一种像水、电设施一样的 IT 基础设施。

根据美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)的定义,云计算是一种利用互联网实现随时随地、按需、便捷地访问共享资源池(如计算设施、存储设备、应用程序等)的计算模式。计算机资源服务化是云计算重要的表现形式,它为用户屏蔽了数据中心管理、大规模数据处理、应用程序部署等问题。

云计算的基本思路十分简单,就是“合”的思路:服务提供商提供应用程序,服务提供商的数据中心负责集中存储过去一直保存在最终用户个人计算机上或企业自己的数据中



心的信息，用户则通过互联网远程访问这些应用程序和数据。

根据上面的定义，“云”就是一些可以自我维护和自我管理的虚拟资源。它通常由一些大规模服务器集群(包括计算服务器、存储服务器、宽带资源等)组成。云计算将所有的资源集中在一起，并由软件自动管理。这使得用户无需为许多的细节而烦恼，能够把更多的精力放在自己的业务上，有利于创新和降低成本。

这种虚拟资源之所以称为“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，云的边界是模糊的；云在空中飘忽不定，无法也无需确定它的具体位置。将其称为“云”，还因为云计算的鼻祖之一——亚马逊(Amazon)公司将其网格计算的产品取名为“弹性计算云(Elastic Compute Cloud, EC2)”，而该产品在商业上取得了很大的成功。

另外，在各种图示中，互联网常用一个云状图案来表示。因此提供资源的网络被称为“云”，同时“云”也是对底层基础设施的一种抽象。“云”的形象描述如图 1.1 所示。

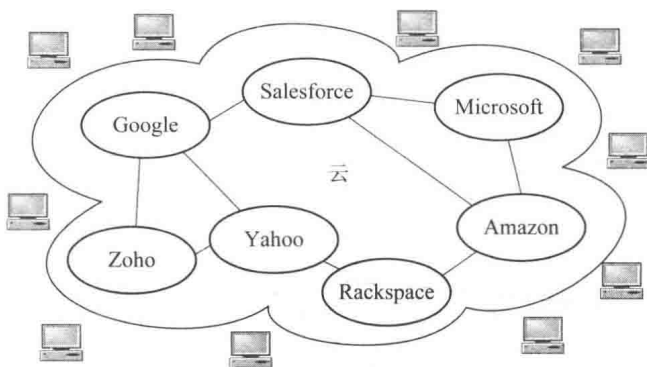


图 1.1 “云”的形象描述

云计算的出现，使提供计算能力的方式发生了巨大变化，这就好比电力供应从古老的单台发电机模式转向电厂集中供电模式。这一改变意味着计算能力和存储能力可以像煤气、水电一样作为商品进行流通，并且取用方便，费用低廉。而其中最大的不同只是在于，计算能力和存储能力是通过互联网进行传输的。

云计算是分布式计算、互联网、大规模资源管理等技术融合与发展的结果。云计算与相关技术的关系如图 1.2 所示。其研究和应用是一个系统工程，涵盖了数据中心管理、资源虚拟化、海量数据处理、计算机安全等重要问题。

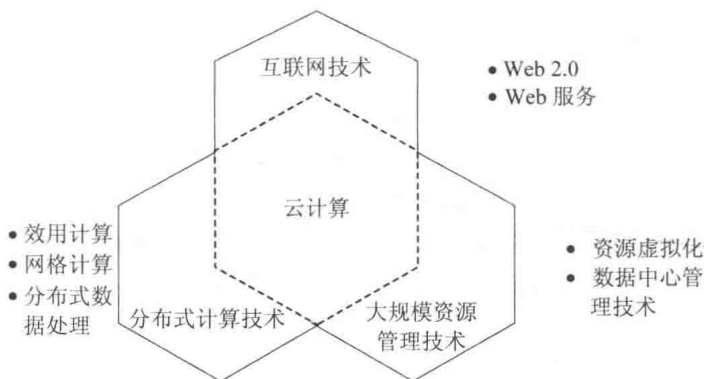


图 1.2 云计算与相关技术的关系



1.1.2 云计算的特征

云计算有以下七大特征：

(1) 超大规模。“云”具有相当大的规模，Google 云计算已经拥有 200 多万台服务器，Amazon、IBM、Microsoft、Yahoo 等的“云”均拥有几十万台服务器。企业私有云一般拥有数百到上千台服务器。“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”，应用在“云”中某处运行，用户无需了解，也不用担心应用运行的具体位置。只需要一台笔记本或者一部手机，就可以通过网络来获取需要的服务，甚至实现超级计算这样的任务。

(3) 高可靠性。“云”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可靠。

(4) 通用性。云计算不针对特定的应用，在“云”的支撑下可以构造出千变万化的应用，同一个“云”可以同时运行不同的应用。

(5) 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。

(6) 按需服务。“云”是一个庞大的资源池，用户可按需购买。使用“云”时，可以像使用自来水、电、煤气那样计费。

(7) 价格低廉。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之传统系统大幅提升，因此用户可以充分享受“云”的低成本优势，通常只要花费几百美元、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。

1.1.3 云计算系统的组成

以电力网络模拟云计算系统的组成，如图 1.3 所示。云计算系统包含以下三个部分：

(1) 以数据中心为代表的信息电厂。

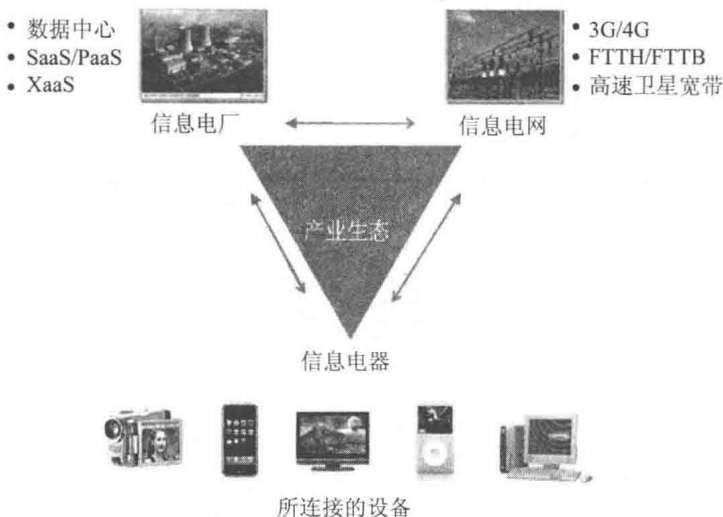


图 1.3 云计算系统的组成



(2) 以手机、电视、计算机等访问终端为代表的信息电器。

(3) 在两者之间起到互联互通作用的信息电网，如移动通信网、光纤网、卫星通信网等。

云计算的访问终端具有丰富多样化的特征。现在用来连接网络的设备丰富多样，除了手机、计算机，还有多媒体座机、网络电台、电子相框、游戏机、音乐播放器等设备。

例如，通过如图 1.4 所示的多媒体座机可以进行网络订票，查看电影介绍。多媒体座机支持音乐播放、Wi-Fi 上网或者宽带拨号上网，还支持 Android 系统及其应用软件。



图 1.4 多媒体座机

1.1.4 云计算的部署模式

云计算可以有三种部署模式，如图 1.5 所示，即公共云、私有云和混合云，三者的比较如表 1.1 所示。

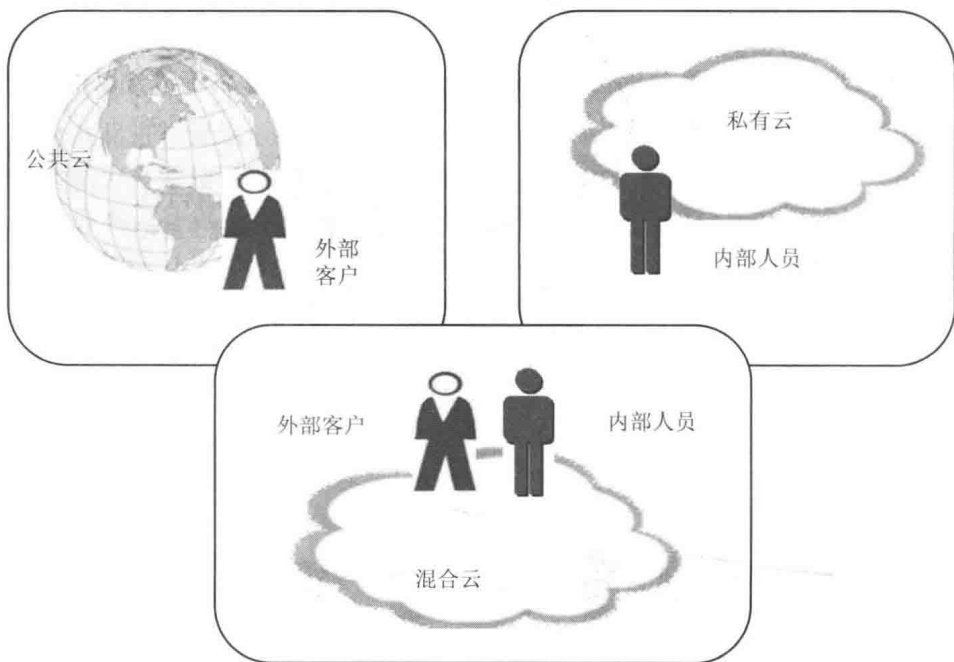


图 1.5 云计算的部署模式



表 1.1 云计算的部署模式的比较

比较项目	前期成本	运行成本	适用对象	安全性	对云资源可用性的控制	法律问题
公共云	无需前期成本	运营可委托服务供应商,按需付费	大企业以及中小企业	低,有更多受到诸如 DDoS(分布式拒绝服务攻击)恶意攻击的可能,需要可信的虚拟数据中心	公司无法自行控制	存在跨国界的数据存储问题
私有云	需要较大的系统投资	需要专门的运营管理技术和开销	大企业	高	公司可自行控制	无跨国界的数据存储问题
混合云	需要一定的系统投资	介于公共云和私有云中间	有多个外部和内部供应商的企业	有应用兼容性问题	公司可在一定程度上自行控制	存在跨国界的数据存储问题

1. 公共云

公共云是指为外部客户提供服务的“云”。它所有的服务是供别人使用的,而不是供自己使用的。目前,典型的公共云有微软的 Windows Azure 平台、亚马逊的 AWS(Amazon Web Services)、Salesforce.com 以及国内的阿里云、用友伟库等。

对于使用者而言,公共云的最大优点是其所应用的程序、服务及相关数据都存放在公共云的提供者处,使用者无需做相应的投资和建设。目前最大的问题是,由于数据不存储在自己的数据中心,其安全性存在一定风险。同时,公共云的可用性不受使用者控制。

2. 私有云

私有云是指企业自己使用的“云”。它所有的服务不是供别人使用的,而是供自己内部人员或分支机构使用的。私有云的部署比较适合于有众多分支机构的大型企业或政府部门。随着这些大型企业数据中心的集中化,私有云将会成为其部署 IT 系统的主流模式。

相对于公共云,私有云部署在企业内部,因此其数据安全性高,系统可用性可自己控制,但其缺点是投资较大。

3. 混合云

混合云是指供自己和客户共同使用的“云”,它所提供的服务既可以供别人使用,也可以供自己使用。相比较而言,混合云的部署方式对提供者的要求较高。

1.2 云计算应用实例

1. 云办公

作为云计算应用实例,本节介绍利用 Google 的云办公应用套件,完成一系列网上应用。

(1) 申请一个基于 Web 的电子邮件账户,账户名为“wadehao.bj@gmail.com”,并通过 Gmail 收发邮件,如图 1.6 所示。图中显示了收件箱内容和邮箱的剩余空间,同时,用户可以通过 Google Talk(聊天)与同事和朋友聊天。