



云计算理论与实践

◎ 潘 虎 著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

云计算理论与实践

潘 虎 著

電子工業出版社·

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书介绍了云计算的基本理论及相关实践,强调理论与实践并重。主要内容包括:云计算技术的基本理论、与其他计算泛型的区别和联系,以及代表性的云计算平台系统;虚拟化技术的基本理论、应用类型及其技术路线;商业虚拟化平台 VMware vSphere、开源虚拟化平台 OpenStack 的系统架构与核心组件,以及详细的安装部署和运维方法;大数据处理技术的基本理论、功能特点与关键技术;开源大数据系统 Hadoop 的系统架构、核心组件与关键技术,及其安装部署和运维方法。

本书内容全面、条理清晰、可操作性强。既注重理论讲解,又注重实践操作。读者通过学习理论知识,结合云平台的实践操作,可对云计算技术快速入门,在此基础上对云计算开展深入研究。为方便教学,本书还配有电子课件及教学资源包,读者可登录华信教育资源网(www.hxedu.com.cn)免费注册下载。

本书既可作为学习云计算理论的技术书籍,又可作为实践技能的培训教材。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

云计算理论与实践 / 潘虎著. —北京:电子工业出版社, 2016.12

ISBN 978-7-121-30194-0

I. ①云… II. ①潘… III. ①云计算—高等学校—教材 IV. ①TP393.027

中国版本图书馆 CIP 数据核字(2016)第 258866 号

策划编辑:戴晨辰

责任编辑:戴晨辰

印 刷:北京京师印务有限公司

装 订:北京京师印务有限公司

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:17.75 字数:454.4 千字

版 次:2016 年 12 月第 1 版

印 次:2016 年 12 月第 1 次印刷

定 价:42.00 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, 88258888。

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

本书咨询联系方式: dcc@phei.com.cn, 192910558(QQ 群)。

前 言

云计算 (Cloud Computing) 是一种动态易扩展的计算方式和信息处理方式。云计算技术的优势在于: 理论上讲, 使得系统的存储能力和计算能力可以无限扩展。云计算思想的产生源于企业对 IT 设施投资及管理的实际需求, 其更关注如何以低成本扩展 IT 系统并使其易于管理。

近几十年来, 计算模式和信息处理模式经历了从大机时代的终端-主机模式 (T-S 模式) 到个人 PC 时代的客户机-服务器模式 (C-S 模式), 再到互联网时代的浏览器-服务器模式 (B-S 模式) 的不断演进。随着物联网和互联网应用需求的增大, 以及移动宽带网络的普及, 用户向互联网输入的数据迅速增长, 软件更多地以服务的形式通过互联网被发布和访问。这些日益增长的业务需要海量存储和强大的计算能力来支撑。云计算的特性满足了现实需求, 迅猛地发展起来。

当前计算机技术发展的广度和深度是以前无法想象的。计算机专业技术人员想要具备更高的技术能力, 就必须快速地掌握各项新技术。多年来, 作者一直专注于云计算平台的搭建、使用和运维管理, 并开展了云计算理论与实践方面的教学工作。在实际工作中体会到优秀技术资料的稀缺, 学习积累过程的不易。为了帮助相关专业人员高效、准确和较为全面地掌握云计算技术的关键知识和技能, 作者结合多年教学经验编写本书, 希望对读者有所帮助。本书中的实践内容均在实验室和实验教学中进行了验证。

本书的主要内容如下。

第 1 章 云计算概述。本章首先阐述了云计算技术的架构, 服务的类型和特点, 以及未来的发展方向。随后从服务层次、部署方式方面介绍了云计算的分类、关键技术, 以及云计算、网格、集群、并行计算、分布式计算、效用计算等计算类型的区别与联系。

第 2 章 虚拟化概述。本章首先阐述了虚拟化技术的定义和分类, 详细讲解了几种重要的服务器虚拟化技术, 并从性能与用户体验等多个方面比较了这几种虚拟化技术的优势与不足, 最后分析了虚拟化技术与云计算的关系。

第 3 章 VMware vSphere 概述。本章首先介绍了 VMware 公司及其产品线的发展, 然后重点介绍了 VMware vSphere 的系统架构、核心组件、功能特点, 最后介绍了 VMware ESXi 服务器的系统架构。

第 4 章 VMware vSphere 平台的搭建和使用。本章详细介绍了该平台的搭建过程, 读者可以仿照虚拟环境下服务器与网络的配置方法, 使用物理机搭建系统平台。

第 5 章 VMware vSphere 配置和高级特性。本章详细介绍了该平台的日常运维方法和高级特性。

第 6 章 OpenStack 概述。本章阐述了开源虚拟化管理平台 OpenStack 的发展历程、系统架构、核心组件、工作流程及生产环境的配置模式, 详细介绍了 OpenStack 组件的功能和结构。

第 7 章 OpenStack 平台的搭建与使用。本章详细介绍了 OpenStack 平台的安装和配置方法。

第 8 章 大数据概述。本章首先阐述了大数据的结构类型和数据特征, 然后介绍了大数据处理的关键技术, 以及大数据处理系统的功能、特性, 其与云计算的关系, 最后介绍了当下流行的大数据系统的实例及大数据处理技术的经典应用。

第9章 Hadoop 大数据技术。本章阐述了开源大数据平台 Hadoop 的系统架构和关键技术，其中主要介绍了 HDFS 和 MapReduce 的工作原理，以及 YARN 框架的工作流程。

第10章 Hadoop 平台的搭建和使用。本章首先详细介绍了 Hadoop 平台的安装、配置、使用及管理方法，然后通过分析 MapReduce 程序，详细阐述了 Hadoop 并行工作的原理和过程。

为方便教学，本书还配有电子课件及教学资源包，读者可登录华信教育资源网（www.hxedu.com.cn）免费注册下载。

作者在本书中使用了部分网络中的资源，如 VMware 公司官网、OpenStack 官网、Hadoop 官网的相关资料，专业技术论坛资源及个人博客文章等。这些网络资源为本书提供了很好的参考素材，在此向相关作者表示衷心的感谢。

在本书的写作过程中，吴钊教授为本书的编写提出了许多宝贵的意见和建议，并参与了本书的框架编写工作，熊伟老师、胡春阳老师和程虹老师为云计算相关理论资料的收集和內容编写、实验平台的搭建做出了大量工作，部分实验的验证工作由鲍贵明、张仕山、龙源等完成。

本书得到了国家自然科学基金（61172084）、湖北省科技支撑计划项目（2013BHE022）、湖北省自然科学基金项目（2013CFC026）、湖北省高校优秀中青年科技创新团队项目（T201413）和湖北文理学院特色教材基金项目的资助。由于作者的水平和时间所限，书中难免存在错误和不妥之处，欢迎读者和业界同仁不吝指正。作者 E-mail: 583436358@qq.com。

作者

2016年12月

目 录

第 1 章 云计算概述	1
1.1 云计算简介	1
1.2 云计算的分类	2
1.2.1 IaaS、PaaS 和 SaaS	2
1.2.2 IaaS 开发和 PaaS 开发	4
1.3 云计算的特点	5
1.3.1 基本特点	5
1.3.2 云计算与其他集群计算	7
1.4 云计算关键技术	9
1.5 云计算应用	11
1.5.1 云计算平台	11
1.5.2 云计算衍生产品	14
1.6 小结	15
深入思考	16
第 2 章 虚拟化概述	17
2.1 虚拟化技术简介	17
2.1.1 计算机虚拟化（服务器虚拟化）	17
2.1.2 存储虚拟化	18
2.1.3 网络虚拟化	20
2.1.4 应用虚拟化	20
2.1.5 桌面虚拟化	21
2.2 服务器虚拟化	22
2.2.1 服务器虚拟化简介	22
2.2.2 服务器虚拟化分类	22
2.2.3 服务器虚拟化用途	24
2.3 x86 虚拟化技术	25
2.3.1 x86 虚拟化技术的发展	25
2.3.2 x86 虚拟化的特征	25
2.3.3 x86 虚拟化技术细节	26
2.3.4 x86 架构服务器虚拟化系统厂商及其产品	28
2.4 KVM 虚拟化技术	31
2.4.1 KVM 简介	31
2.4.2 KVM 的 CPU 虚拟化	32
2.4.3 KVM 用户物理内存管理	34

2.5	云计算与虚拟化	35
2.5.1	部署与应用	35
2.5.2	向服务转型	35
2.6	小结	36
	深入思考	37
第 3 章	VMware vSphere 概述	38
3.1	VMware 公司简介	38
3.1.1	x86 系统虚拟化技术的提出	38
3.1.2	VMware 公司发展历史	39
3.2	VMware 公司产品概述	40
3.2.1	核心产品设计理念	40
3.2.2	产品简介	40
3.3	VMware vSphere 组成与功能	42
3.3.1	VMware vSphere 主要组件	42
3.3.2	VMware vSphere 基本功能	43
3.3.3	VMware vSphere 高级功能	44
3.3.4	VMware vSphere 插件	45
3.4	VMware vSphere 逻辑分层结构和物理拓扑	46
3.4.1	VMware vSphere 虚拟化层	46
3.4.2	VMware vSphere 管理层	49
3.4.3	VMware vSphere 接口层	49
3.4.4	VMware vSphere 数据中心的物理拓扑	50
3.5	VMware ESXi 架构	51
3.5.1	Service Console	51
3.5.2	VMkernel	52
3.5.3	ESXi 小结	54
3.6	VMware vSphere 5.5 特点	55
3.7	VMware vSphere 存储	55
3.7.1	硬盘分类	55
3.7.2	磁盘阵列	57
3.7.3	存储分类	57
3.8	小结	59
	深入思考	60
第 4 章	VMware vSphere 平台的搭建和使用	61
4.1	ESXi 服务器的安装和配置	62
4.1.1	安装 ESXi 服务器	62
4.1.2	配置 ESXi 服务器网络	67
4.2	安装配置 Openfiler 服务器	71
4.2.1	安装 Openfiler 虚拟机	71

4.2.2	配置 Openfiler 虚拟机	78
4.3	安装并配置 VMware vCenter	84
4.3.1	安装 Windows Server 2003	85
4.3.2	安装虚拟光驱	86
4.3.3	安装 vCenter Single Sign On	87
4.3.4	安装 VMware vCenter Inventory Service	91
4.3.5	安装 VMware vCenter Server	93
4.4	登录 vCenter 并挂载 ESXi 主机	99
4.5	连接 Openfiler 存储	101
4.6	创建虚拟机	107
4.7	小结	112
	深入思考	113
第 5 章	VMware vSphere 配置和高级特性	114
5.1	修改硬件参数	114
5.2	查看虚拟机文件	115
5.3	快照的使用	116
5.4	虚拟机转模板	119
5.5	虚拟机迁移	122
5.6	分布式资源调配 DRS	125
5.6.1	创建 DRS 群集	126
5.6.2	体验 DRS	129
5.7	资源池的使用	133
5.8	虚拟机的高可用性	134
5.9	热备功能	137
5.10	虚拟网络	138
5.11	存储网络	140
5.12	小结	140
	深入思考	141
第 6 章	OpenStack 概述	142
6.1	OpenStack 简介	142
6.1.1	OpenStack 与云计算	142
6.1.2	OpenStack 的功能	143
6.1.3	OpenStack 的发展历程	143
6.1.4	KVM 开放虚拟化技术	144
6.2	OpenStack 架构	145
6.3	OpenStack 工作流程	146
6.3.1	Bexar 版本的工作流程	146
6.3.2	Folsom 版本的工作流程	149
6.4	OpenStack 生产环境的配置模式	150

6.5	OpenStack 各组件详解	151
6.5.1	Nova 组件	151
6.5.2	Keystone 组件	151
6.5.3	Neutron 组件	153
6.5.4	Swift 组件	154
6.5.5	Cinder 组件	157
6.5.6	Glance 组件	158
6.5.7	Horizon 组件	159
6.5.8	Heat 组件	160
6.6	OpenStack 在企业中的应用	160
6.6.1	小米 OpenStack 项目概况	160
6.6.2	联想 OpenStack 的高可用企业云平台实践	161
6.6.3	OpenStack 在天河二号的大规模部署实践	162
6.7	VMware 与 OpenStack 的比较	162
6.8	小结	164
	深入思考	164
第 7 章	OpenStack 平台的搭建与使用	165
7.1	实验环境资源需求	165
7.2	实验环境拓扑	165
7.3	实验环境配置	167
7.4	安装和配置 Identity Service (身份服务)	172
7.4.1	先决条件	172
7.4.2	安装并配置组件	172
7.4.3	配置 Apache HTTP 服务	173
7.4.4	完成安装	174
7.4.5	创建临时管理员令牌环境	174
7.4.6	创建服务实体和 API 端点	175
7.4.7	创建域、项目、用户和角色	175
7.4.8	验证操作	177
7.4.9	创建脚本	178
7.4.10	使用脚本	179
7.5	安装和配置 Image Service (映像服务)	179
7.5.1	先决条件	179
7.5.2	安装和配置组件	181
7.5.3	完成安装	182
7.5.4	确认安装	183
7.6	安装和配置 Compute Service (计算服务)	183
7.6.1	安装并配置管理节点	183
7.6.2	安装和配置计算节点	187

7.7	安装配置 Networking Service（网络服务）	190
7.7.1	安装和配置管理节点	191
7.7.2	安装和配置计算节点	197
7.8	安装和配置 Dashboard	203
7.8.1	安装和配置组件	203
7.8.2	完成安装	204
7.8.3	验证操作	204
7.9	安装和配置 Block Storage Service（块存储服务）	204
7.9.1	安装和配置管理节点	204
7.9.2	安装和配置一个存储节点	208
7.10	Horizon 操作	211
7.11	自动化部署	215
7.12	小结	216
	深入思考	217
第 8 章	大数据概述	218
8.1	大数据简介	218
8.1.1	大数据的定义	218
8.1.2	大数据的结构类型	218
8.1.3	大数据的特征	219
8.1.4	大数据的处理技术	219
8.2	大数据处理系统	222
8.2.1	大数据处理系统的功能	222
8.2.2	大数据处理系统的特性	222
8.2.3	云计算与大数据处理系统	223
8.3	大数据处理系统实例	223
8.3.1	Google 大数据处理系统	223
8.3.2	Hadoop	225
8.4	大数据应用	226
8.4.1	精准广告投放	226
8.4.2	精密医疗卫生体系	227
8.4.3	个性化教育	227
8.4.4	交通行为预测	227
8.4.5	数据安全	228
8.5	小结	228
	深入思考	228
第 9 章	Hadoop 大数据技术	229
9.1	Hadoop 概述	229
9.1.1	Hadoop 简介	229
9.1.2	Hadoop 编年史	230

9.1.3	Hadoop 架构	233
9.1.4	Hadoop 组件	234
9.2	HDFS 概述	235
9.2.1	HDFS 简介	235
9.2.2	HDFS 工作特性	237
9.2.3	文件读取过程	241
9.2.4	文件写入过程	241
9.3	MapReduce 工作原理	242
9.4	Shuffle 过程	244
9.4.1	Map 端	244
9.4.2	Reduce 端	245
9.5	YARN 架构任务调度	246
9.5.1	MRv2 (MapReduce Version 2) 的基本组成	246
9.5.2	YARN 的基本组成	246
9.5.3	YARN 架构下 MapReduce 的任务流程	247
9.6	Hadoop 应用领域	248
9.7	小结	249
	深入思考	250
第 10 章	Hadoop 平台的搭建和使用	251
10.1	Linux 系统配置	251
10.2	Hadoop 配置部署	257
10.3	运行 Hadoop	260
10.4	HDFS Shell 命令操作	263
10.5	MapReduce 程序解读	266
10.6	小结	270
	深入思考	270
	参考文献	271

云计算概述

1.1 云计算简介

云计算（Cloud Computing）是基于互联网的相关服务的增加、使用和交付模式，云计算通过互联网提供动态易扩展且经常是虚拟化的资源。美国国家标准与技术研究院（NIST）这样定义了云计算：云计算是一种按使用量付费的模式，这种模式提供可用的、便捷的、按需的网络访问，进入可配置的计算资源共享池（资源包括网络、服务器、存储、应用软件、服务），这些资源能够被快速提供，而只需投入很少的管理工作，或服务供应商进行很少的交互。

计算机技术、计算模式和信息处理模式随着用户信息处理需求的增长而产生和发展。从计算机出现至今，计算模式和信息处理模式经历了大机时代的终端-主机模式（T-S 模式），PC 时代的客户机-服务器模式（C-S 模式），互联网时代的浏览器-服务器模式（B-S 模式）。目前，随着物联网和互联网应用需求的增大及移动宽带网络的普及，用户向互联网输入的数据量迅速增加。同时，软件多以服务的形式通过互联网发布和访问。这些日益增长的业务需要海量的存储和强大计算能力来支持。为了满足不断增长的业务需要，如果还沿用传统的数据中心架构，付出的软硬件成本会大大增加。云计算是一种动态、易扩展的计算方式和信息处理方式，存储能力和计算能力理论上可以无限增大，满足信息处理等业务需求的快速增长。而且云计算使用 x86 架构的服务器，性价比很高。这些优点使云计算逐渐发展和流行起来。云计算的原理是：使用特定的软件，按照指定的优先级和调度算法，将数据计算和数据存储分配到云计算集群中的各个节点计算机上，节点计算机并行运算，处理存储在本节点上的数据，结果回收后合并。云计算的产生并非来自学术界，而是产生于企业计算和互联网领域，它是分布式计算（Distributed Computing）、并行计算（Parallel Computing）、效用计算（Utility Computing）、网络存储（Network Storage Technologies）、虚拟化（Virtualization）、负载均衡（Load Balance）、热备份冗余（Hot Backup Redundancy）等传统计算机和网络技术发展融合的产物。云计算更关心如何扩展系统，如何方便 IT 管理，如何降低成本。

云计算通过互联网向用户提供服务，这些服务包括运算服务，例如，希望通过海量的销售记录计算某个大型商业网站某类商品最近几年的销售量，用户向云服务前端提交任务，由“云”返回计算结果；基础设施服务，例如，用户向云服务前端申请一台服务器，指明自己对硬件和软件的需求，包括 CPU 需求，使用多大内存和硬盘，操作系统是什么等，“云”将按照用户的要求虚拟一台服务器供使用，登录服务器（使用远程桌面或终端工具软件登录），会发现服务器的配置与用户的要求一致。当然还有许多其他的服务类型，比如云存储、云安全等。

对于用户来说,只需向“云”提出要求得到服务,不需要了解云内部的细节。这里的“云”实际上是一个大量硬件和软件的集合体,这些软硬件集合通过网络和“云软件”连接和组织在一起,向用户提供各种服务。前面提到的虚拟服务器,CPU 和内存来源于哪里,销售量运算究竟是哪几台机器做的,用户并不需要知道,而是由“云软件”组织调配“云”中的资源完成。“云软件”可以看作云资源集合的操作系统,有着操作系统的特征:管理软硬件资源和任务流程,提供人机界面。在需要时,可以向这个集合体内增加软硬件资源,不需要时可以把软硬件资源从这个集合体分离出去。

综上所述,云计算可以看成是一种 IT 资源的交付和使用模式,用户通过网络,以按需、易扩展的方式获得所需的资源(包括硬件、平台和软件)。“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、随时扩展、按需付费。这种特性被人们形象地称为像使用水电一样使用 IT 资源。计算能力也可以作为一种商品进行流通,就像水电一样,取用方便,通过互联网进行传输。之所以称为云计算,是因为在计算机网络拓扑图中互联网通常以云表示,在互联网之中的云资源集合也可以抽象为一朵云。用户对云资源的需求往往是根据业务的实际需求来衡量的,需要多少资源就使用多少,使用完成后再还给“云”,由“云”再提供给其他用户使用。

云计算未来主要有两个发展方向:一是构建与应用程序紧密结合的大规模底层基础设施,使得应用能够扩展到很大的规模;二是通过构建新型的云计算应用程序,在网络上提供更加丰富的用户体验。云计算虽然是一种新型的计算模式,但是现实的需要恰恰为云计算提供了良好的发展机遇。虽然现在的云计算并不能完美地解决所有的问题,但是相信在不久的将来,一定会有更多的云计算系统投入使用,云计算系统也将不断地被完善,并推动其他科学技术的发展。

1.2 云计算的分类

1.2.1 IaaS、PaaS 和 SaaS

按照提供服务的层次和类别,云计算可分为 3 类: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)。“基础设施”是一个应用系统的硬件平台,处于应用系统的下端;“平台”包括操作系统,中间件和函数库;“软件”就是整个应用系统。在不同的服务类型下,用户可以控制的内容和云服务控制的内容有所区别,如图 1.1 所示。

1. IaaS

IaaS 为用户提供了计算基础架构,通常指提供了物理机、虚拟机、网络资源及其他资源(如虚拟机映像库、块存储或者基于文件的存储、防火墙、负载均衡、IP 地址、虚拟局域网等)。以前,如果要在企业平台运行一些企业应用,需要购买服务器,以及其他价格高昂的硬件来支持。现在,用户可以使用 IaaS 方式将硬件外包,IaaS 公司将提供虚拟服务器、存储和网络硬件,用户可以租用来运行企业应用,有效地节省了维护成本和办公场地。用户可以在任何时候利用租用来的硬件运行其应用。国际上主要的 IaaS 提供商和产品包括:亚马逊的 AWS、

微软的 Azure、Rackspace 的 OpenStack、IBM 的 SoftLayer、VMware 的 vCloud 等。国内的有阿里云、青云 (Qing Cloud) 及中国移动的大云 (Big Cloud) 等。本书后续章节所讲述的 VMware vSphere 和 OpenStack 是两种有代表性的 IaaS 平台软件, 它们都是采用虚拟技术, 提供虚拟机 IaaS 服务。

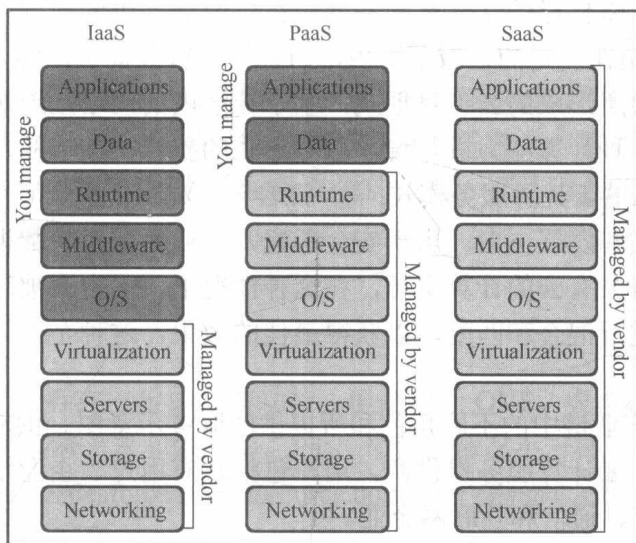


图 1.1 云计算的分类及可控内容

2. PaaS

PaaS 也称为中间件服务。为用户提供的服务平台通常包括操作系统、编程语言运行环境、数据库和大数据集处理、Web 服务器等。PaaS 把开发环境作为一种服务来提供, 可以使用中间商的设备和软件开发自己的程序, 通过互联网和服务器传送给用户。软件公司的软件开发和软件运行都可以在这一层进行, 节省了时间和资源。PaaS 公司在网上提供各种开发和分发应用的解决方案, 开发出的应用调用了 PaaS 平台的 API, 运行时使用了 PaaS 平台软硬件, 节省了硬件开销, 也使得处在分散工作场地之间的合作变得更加容易。PaaS 服务提供网页应用管理、应用设计、应用虚拟主机、存储、安全及应用开发协作工具等。主要的服务平台包括亚马逊的 AWS Elastic Beanstalk、Heroku、Force.com、谷歌的 GAE (Google App Engine)、新浪的 SAE、百度云的开发引擎等。

大数据处理系统是一种 PaaS 平台。初期, Google 的创始者因为无法购买昂贵的商用服务器运行搜索引擎, 于是采用了大量廉价的 x86 架构个人计算机组成集群来提供搜索服务, 成功地把这种个人计算机集群的运算能力做到比商用服务器更强大, 而成本却远远低于商用的硬件与软件, 从而产生了大数据处理技术。2003—2006 年, Google 发表了 4 篇关于分布式文件系统、并行计算、数据管理和分布式资源管理的文章, 奠定了大数据处理技术发展的基础。基于这些文章, 开源软件 Hadoop 逐步复制了 Google 的云计算系统, 从此 Hadoop 大数据处理系统平台开始流行。本书后续章节将介绍 Hadoop 平台的结构和使用。

3. SaaS

SaaS 为用户提供了按需支付费用 (On-Demand) 的应用软件。用户不必操心各种应用程序的安装、设置和运行维护, 一切都由 SaaS 服务提供商来完成。用户只需要支付费用, 通过可视

化的客户端来使用它，如谷歌的 Apps、微软的 Office 365、Citrix 的 CloudStack，以及目前流行的各种云存储（网盘）、云相册、云备份、云打印、云监控等针对个人使用的云服务产品。

云计算的部署方式有 4 类，包括：① 私有云，数据中心部署在企业内部，由企业自行管理；② 公共云，数据中心由第三方的云计算供应商提供，供应商帮助企业管理基础设施（如硬件、网络等），企业将自己的软件及服务部属在供应商提供的数据中心，并且支付一定的租金；③ 社区云，是指在一定的地域范围内，由云计算服务提供商统一提供计算资源、网络资源、软件和服务能力所形成的云计算形式，基于社区内的网络互连优势和技术易于整合等特点，通过对区域内各种计算能力进行统一服务形式的整合，结合社区内的用户需求共性，实现面向区域用户需求的云计算服务模式；④ 混合云，混合云融合了公有云和私有云，是近年来云计算的主要模式和发展方向。出于安全考虑，企业更愿意将数据存放在私有云中，但是同时又希望可以获得公有云的计算资源，在这种情况下，混合云被越来越多地采用。混合云将公有云和私有云进行混合和匹配，以获得最佳的效果，这种个性化的解决方案，达到了既省钱又安全的目的。

图 1.2 是云计算产业链中的不同角色和所提供的服务示意图。供应商包括基础设施制造商、基础设施运营商、云计算服务提供商。用户包括政府用户、企业用户和个人用户，行业又分政府、教育、医疗、通信和互联网企业。

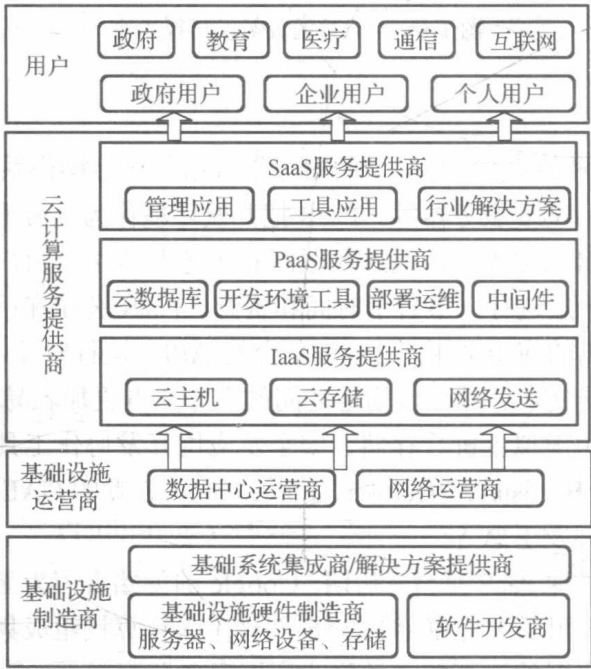


图 1.2 云计算产业链

1.2.2 IaaS 开发和 PaaS 开发

PaaS 是将一个开发和运行的平台作为服务提供给用户，PaaS 开发是在该平台上开发具体业务应用；IaaS 是将虚拟机或其他资源作为服务提供给用户，IaaS 开发是开发资源组织和发布平台，目前国内的大量云平台公司开发虚拟机管理平台、桌面云平台，几乎都是基于 KVM 技术，在 OpenStack 基础上开发的。以下从几个方面比较两者的区别。

1. 开发环境

PaaS 提供商都会给开发者提供一整套包括 IDE 在内的开发和测试环境。IaaS 开发人员主要沿用之前的开发环境（如 Eclipse），由于开发环境与云的整合比较欠缺，所以在使用时不是很方便。

2. 支持的应用

因为 IaaS 主要提供虚拟机，且普通的虚拟机支持多种操作系统，所以 IaaS 支持的应用范围非常广泛。但如果一个应用运行在某个 PaaS 平台上，则不是一件轻松的事，因为不仅需要确保这个应用是基于这个平台所支持的语言，而且也要确保这个应用只能调用此平台所支持的 API，如果应用调用了平台所不支持的 API，那么就需要对这个应用进行修改。

3. 开放标准

虽然很多 IaaS 平台都存在一定的私有功能，但是由于 OVF（虚拟机文件格式）等协议的存在，使得 IaaS 开发可以跨平台和供应商。PaaS 平台的情况不一样，无论是 Google 公司的 Google App Engine，还是 Salesforce 公司的 Force.com 都存在私有 API，应用不能跨平台。

4. 可伸缩性

PaaS 平台会自动调整资源来帮助运行于其上的应用更好地应对突发流量。IaaS 平台则需要开发人员手动对资源进行调整才能应对。

5. 整合率和经济性

PaaS 平台整合率非常高，如 PaaS 的代表 Google App Engine 能在一台服务器上承载成千上万的应用。而普通的 IaaS 平台的整合率最多不会超过 100 个，普遍在 10 个左右，使得 IaaS 的经济性不如 PaaS。

6. 计费和监管

PaaS 平台在计费和监管方面不仅达到了 IaaS 平台的操作系统层面（如 CPU 和内存的使用量），而且还能达到应用层面（如应用的反应时间或应用所消耗的事务多少等），这将提高计费和管理的精准性。

7. 学习难度

在 IaaS 上开发和管理应用的方式与常规方式比较接近，而在 PaaS 上开发则有可能需要学习一门新的语言或新的框架，所以 IaaS 学习难度相对低。

1.3 云计算的特点

1.3.1 基本特点

云计算运行在“云”上，“云”是一个由大量的硬件和软件组成的集合体，硬件通常指一个由高速网络连接在一起的计算机集群，云软件组织调配资源，提供图形化界面或 API 接口。

1. 超大规模

“云”具有超大的规模，Google 云计算拥有 100 多万台服务器，Amazon、IBM、微软、Yahoo 等的“云”也拥有几十万台服务器。一般大型企业的私有云也拥有数百台服务器。超大规模的计算机集群能赋予用户前所未有的计算能力。

2. 虚拟化

虚拟化包括资源虚拟化和应用虚拟化。资源虚拟化是指异构硬件在用户面前表现为统一资源；应用虚拟化是指应用部署的环境和物理平台无关，通过虚拟平台对应用进行扩展、迁移、备份，这些操作都是通过虚拟化层完成的，虚拟化技术支持用户在任意位置、使用各种终端获取应用服务，如大数据处理系统。使用虚拟化技术，用户所请求的资源来自“云”，应用在“云”中运行，用户无须了解，也不用担心应用运行的具体位置。只需要一台笔记本或一部手机，就可以通过网络服务实现用户需求，甚至包括超级计算这样的任务。

3. 动态可扩展

云计算能迅速、弹性地提供服务。服务使用的资源能快速扩展和快速释放。对用户来说，可在任何时间购买任何数量的资源。资源可以是计算资源、存储资源和网络资源等。与资源节点相对应的也有计算节点、存储节点和网络节点。如果所需资源无法达到用户需求，可通过动态扩展资源节点增加资源以满足需求。当资源冗余时，可以添加、删除、修改云计算环境的资源节点。冗余可以保证在任一资源节点异常宕机时，不会导致云环境中业务的中断，也不会导致用户数据的丢失。资源动态流转意味着在云计算平台下实现资源调度机制，资源可以流转到的地方。例如，在应用系统业务整体升高的情况下，可以启动闲置资源加入云计算平台中，提高整个云平台的承载能力以应付系统业务的升高。在整个应用系统业务负载低的情况下，可以将业务集中起来，将闲置下来的资源转入节能模式，提高部分资源利用率，以节省能源。

4. 按需部署

供应商的资源保持高可用和高就绪的状态，用户可以按需自助获得资源。按需分配是云计算平台支持资源动态流转的外部特征表现。云计算平台通过虚拟分拆技术，可以实现计算资源的同构化和可度量化，可以提供小到一台计算机，多到千台计算机的计算能力。按量计费源于效用计算，在云计算平台实现按需分配后，按量计费也成为云计算平台向外提供服务时的有效收费形式。

5. 高灵活性

现在大部分的软件和硬件都支持虚拟化，各种 IT 资源（如软件、硬件、操作系统、存储网络等）通过虚拟化放置在云计算虚拟资源池中进行统一管理。云计算能够兼容不同硬件厂商的产品，兼容低配置机器和外设，获得高性能计算。

6. 高可靠性

云计算平台把用户的应用和计算分布在不同的物理服务器上，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，即使单点服务器崩溃，仍然可以通过动态扩展功能部署新的服务器，增加各项资源容量，保证应用和计算的正常运转。