



亚马逊AWS 云基础与实战

王毅 编著

清华大学出版社

亚马逊 AWS 云基础与实战

王毅 编著

清华大学出版社

北 京

内 容 简 介

本书详细介绍了亚马逊 AWS 云服务特性、适用场景及操作方法等,通过列举大量详细案例,旨在让读者全面了解如何利用亚马逊云计算平台完成客户的业务需求和 IT 需求;如何在数分钟内建立属于自己的虚拟数据中心;如何便捷地利用数台到数千台服务器和服务完成传统机房无法想象完成的任务。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

亚马逊AWS云基础与实战 / 王毅 编著. —北京:清华大学出版社, 2017

ISBN 978-7-302-47964-2

I. ①亚… II. ①王… III. ①云计算 IV. ①TP393.027

中国版本图书馆CIP数据核字(2017)第196094号

责任编辑:王 军 李维杰

封面设计:周晓亮

版式设计:思创景点

责任校对:牛艳敏

责任印制:王静怡

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京富博印刷有限公司

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185mm×230mm

印 张:15.75

字 数:313 千字

版 次:2017 年 8 月第 1 版

印 次:2017 年 8 月第 1 次印刷

印 数:1~4000

定 价:49.80 元

产品编号:075233-01



王毅

国内首批获得AWS专家级认证资质人员之一，是最早进入云计算领域的资深从业者，现任YOU WORLD (由我) CTO，曾在亚马逊AWS任大中华区资深解决方案架构师、高级区域主管，还曾在阿里云担任资深技术专家、在IBM全球服务部门担任资深架构师经理，精通亚马逊AWS、阿里云、微软AZURE，帮助许多中国顶级互联网公司，利用云平台的优势，进行技术和商业模式创新。

主要内容

本书详细介绍了亚马逊AWS云服务特性、适用场景及操作方法等，通过列举大量详细案例，旨在让读者全面了解如何利用亚马逊云计算平台完成客户的业务需求和IT需求；如何在数分钟内建立属于自己的虚拟数据中心；如何便捷地利用数台到数千台服务器和服务完成传统机房无法想象完成的任务。

读者对象

本书介绍了你一定用得到的几项最重要的AWS服务。不管你是想要技术型创业，还是想要研究后端技术；不管你想要做机器学习，还是想做大数据研究。本书都可以成为你良好的起点。对于资深技术人员，本书也是一本很好的日常使用的参考书。



推荐序一

“云已经成为信息技术的新常态。”

——安迪·杰西

云计算如今已尽人皆知、无人不晓，成为继个人电脑和互联网之后最重要的信息技术革命。目前对于云计算有着各种不同的定义和描述。简而言之，云计算采用按需使用、按使用付费的方式，通过网络提供各类计算资源。云计算的终极理想是像水电等公共事业一样，将计算资源通过网络高效、灵活、方便地提供给广大用户。

回顾云计算过去十年迅猛发展的历史，我们看到大约相隔5年的三个关键时间点——2006年、2010年和2015年。

在2006年前后，我正忙于推动以基于面向服务的架构(SOA)来构建新一代企业级应用软件。同年夏天，美国IT业有人正式提出了云计算的概念，而亚马逊已经将云计算的商业模式付诸实践，经过一年多的努力，正式推出了首家商业云服务——亚马逊云服务 Amazon Web

Service(AWS)。2006年3月14日,亚马逊云服务率先推出了简单存储服务 Simple Storage Service(S3),而同年秋天亚马逊云服务推出了弹性计算云服务 Elastic Computing Cloud(EC2)。

自2008年云计算的概念开始进入中国。到了2010年,云计算在全球已经取得令人瞩目的成功。国际上各个IT公司纷纷将自己的产品服务贴上云的标签,高高举起了云计算的大旗。在中国,各种名目的云计算大会相继召开,一些云基地、云计算中心开始相继建成,大家开始谈论如何搭上云的快车,通过创新推动经济的发展。当然,由于众说纷“云”,不少人觉得云遮雾罩,不得要领;而亚马逊云服务异军突起,连续发布了一系列关键的计算、存储、网络、数据库、分析等核心云服务,并被创新型企业广泛采用,逐渐成为业内公认的、领先的云服务供应商。

到了2015年,正当国内一些人以为云计算的概念已经红到了头,转而寻找下一个值得炒作的IT概念时,人们却惊奇地注意到云计算已经成为IT的新常态,这个颠覆性的新技术和新服务模式的有机结合已显示出强大的生命力,开始深刻改变IT世界。许多企业明确了云优先的战略,企业从直接云构建云部署应用,直到全部上云(All in on Cloud)。2015年,云计算在中国的应用逐步落地,各类IT企业全力发力,领先的企业开始踏上云的征程。

云计算是信息技术的重大创新,其成功有着必然性,因为它整合了IT各个方面最先进和最合理的元素。技术上,云计算广泛采纳了互联网时代IT技术最新发展的许多成果,如虚拟化、软件定义一切、面向服务的架构、分布式计算和存储、网格和并行计算、海量数据管理、容错技术等。实践上,云计算既广泛地吸取、采纳了软件系统和架构的最佳技术,得力于硬件设备高性能、低成本的设计和制造,又从互联网大规模、大数据、多用户、低成本的数据中心运营维护中吸取了大量经验。业务上,云计算从开源技术、应用的松耦合、无状态到混搭(Mashup)汲取营养,从托管的服务进一步发展成按需服务、按使用付费的商业模式。

云计算带来的好处是多方面的,影响也是广泛和深远的。人们普遍认识到云计算可以有效整合计算资源,避免无法精确预先规划、计算容量的困境,实现弹性的、可扩展的、高可用的服务,通过资源共享和规模效益有效降低计算资源的成本。云计算帮助IT获得敏捷性,打造创新基因,重获创新能力,将IT固定资产投资转为可变运营成本支出,节省IT的总体拥有成本,将IT人员从烦琐的无差异化的劳动中解脱出来,专注于快速提供创新型产品和服务,更好地满足客户的需求。亚马逊云计算除了具备上述优点外,更加充分地发挥了云的可靠性、高可扩展性和高可用性,通过快速迭代和持续创新,使AWS云服务具有深度和广度,有效灵活地降低IT成本,为客户提供优质的服务。

云计算一经推出,就得到软件开发人员和系统运维人员的青睐。云计算意味着没有先期软硬件购置投入,不需要购买或租赁机房设施,也不需要雇用众多的专业技术人员。因此,



云计算很快成为创业公司的宠儿，它们直接在云上开发、部署、运营，很快就取得了骄人的成绩。云计算帮助创业人员以很低的成本使用最新的 IT 技术，让新型的移动互联网企业能快速尝试新的商业模式，搭建可扩展、高可用的社交互联应用。许多创新型企业，依靠数人的 IT 团队，支撑起上亿美金的互联网收入。

云计算在企业级应用方面也取得了长足的进步，大中型企业纷纷开始拥抱云计算。云计算成为以创新促发展、实现“互联网+”改造、完成企业数字化转型的利器。越来越多的企业开始踏上云的征程，在云上实现从测试开发、网站和移动 APP、存储、备份和归档、灾难恢复、大数据分析、高性能计算到企业资源管理、关键业务应用、数据中心向云迁移，直到全部上云。传统企业借助云计算，重新焕发出青春活力。

如今云计算已经渗透到我们日常生活的方方面面，并开始广泛深入地影响和改善人们的生活。当你下载移动 APP、查看地图、上传社交照片、收听网络音乐、观看在线视频、购买电商产品、进行各类网上交易、玩网上游戏时，你可能已经在使用各种各样的云服务。下一阶段可以预期，云计算将与大数据、物联网、社交网络媒体、人工智能和虚拟现实等新兴信息技术交融互动，继续为信息产业带来深刻变革，塑造技术和商业的未来世界。

本书作者中有一些是富有经验的解决方案架构师，有一些是培训部门的资深讲师。他们此前已有 IT 行业多年的从业经验，来到 AWS 后通过系统培训，自学应用和上云实践，掌握了 AWS 有关的知识技能，并在面向客户的架构设计和教学实践中积累了丰富的经验，成为国内最早系统、深入掌握 AWS 的技术专家，也是 AWS 的忠实粉丝。

本书阐述了 AWS 云服务的概念，介绍了云服务的知识和应用。作为国内专门介绍亚马逊云计算服务的图书，本书既可以作为教学和自学 AWS 的简明教材，也可作为云计算从业人员的参考资料。本书适合云计算管理和业务人员、云计算研发人员、云计算架构师和咨询师、云计算开发和运营维护人员，以及云计算应用设计人员学习和参考。结合亚马逊网站上提供的技术资料，本书的读者将能够对亚马逊云服务架构体系有一个比较全面和深入的了解。

如今云计算已不再只是热点概念，而已成为信息技术的重大创新和颠覆性技术。云的时代已经来临，云计算却仍方兴未艾。开卷有益，希望此书助您畅游云的世界。

张侠 亚马逊 AWS 首席云企业顾问



推荐序二

曾几何时，初创公司搭建一台服务公众的服务器是如此困难。首先需要花钱买一台服务器，然后找托管机房，签完各种合同，交完押金和钱，将庞大的机器搬到机房里的某个机柜，插好各种线，然后才可以接入互联网开始服务。此后有了服务器租赁业务，但依然比较麻烦，办完各种手续之后，大概需要两三天，机器才能准备好。在如今云服务提供的便利之下，5分钟就可以开一台服务器，随时随地，用完即关。

如果将传统服务器比作每天早上下山挑水喝，那么云服务就像是自来水入户，打开水龙头，水就流了出来，不用的时候，可以随时关上水龙头。既充分利用了资源，也没有造成资源的浪费，把成本降到了最低。将计算能力分割成块，然后当成自来水一样出售给大众，这无疑是大势所趋。未来，任何一名普通学生想起一个复杂的算法，都可以经过简单的几步操作，在家中开启几百台服务器进行计算，最后在验证完算法的正确性之后关闭它们。云服务会真正像自来水一样走进千家万户。

在众多云服务之中，亚马逊 AWS 是其中的翘楚。其服务的稳定性、易用性，以及 API



的多样丰富，使其占据业界统治地位，遥遥领先于其他竞争对手。一方面，互联网电商亚马逊借助 AWS 完成了技术转型，成为华尔街最看好的科技公司之一；另一方面，AWS 也帮助千万家互联网初创企业成功创业，可以借助低成本云服务迅速开展业务，把握住转瞬即逝的商机。在 AWS 服务推出很多年之后，亚马逊依然推陈出新，不断完善其业务线，不断主动降低价格，可谓诚意十足，业界良心。

木瓜移动从创业之初的 2008 年就在使用亚马逊 AWS 云服务。2014 年 AWS 入华之时，更是很荣幸成为第一批用上中国 AWS 服务的公司之一。在长时间的合作中，AWS 不断地给公司提供良好的服务和帮助，这也是公司成长和壮大的助力之一。

本书介绍了你一定用得到的几项最重要的 AWS 服务。不管你是想要技术型创业，还是想要研究后端技术；不管你想要做机器学习，还是想做大数据研究。本书都可以成为你良好的起点。作为资深技术人员，本书也是一本很好的日常使用的参考书。

钱文杰 木瓜移动联合创始人及 CTO



序

2007 年一次偶然的机会，我接触到了云计算，一下子就被其简捷、方便、强大的特性所吸引，开始疯狂地查找相关资料。当时，有关云计算的著作，国内几乎没有，辗转找到一些国外的书籍，我便如饥似渴地阅读起来。随着我对云计算了解的深入，我越发感觉它的前景远大。果然不出所料，短短几年，云计算已经成为整个行业的大势所趋，甚至可以说，每个人的生活都与之有密不可分的联系。

虽然近年来，国内市场上有关云计算的专业书籍层出不穷，但具体指导用户使用云计算的书籍却凤毛麟角。本书通过分享大量的案例和方法，旨在抛砖引玉，让更多的人使用的过程中领略亚马逊云计算的独特魅力。

本书对亚马逊 AWS 做了初步介绍，并分门别类地列举了亚马逊 AWS 提供的服务及具体操作方法，适用于云计算开发、使用和运维人员，能对读者起到一定的指导作用。

从动念写这本书，到真正完工，历时两年，期间遇到不少困难，再加上亚马逊 AWS 的功能在不断地推陈出新，也就不得不随之不断修改，好在有同事、朋友和家人的支持和鼓励，

我最终还是交出了一份答卷。由于水平有限，不足之处，在所难免。

借此机会，感谢亚马逊 AWS 全球副总裁、中国执行董事容永康先生对本书的大力支持。同时，感谢亚马逊 AWS 中国架构师团队负责人张荣典、孙素梅、曹玮祺和培训团队资深讲师包光磊、张波、黄涛审阅本书以及提出的宝贵意见。同时，感谢亚马逊 AWS 中国资深架构师陈林涛、郑进佳的大力协助。感谢前亚马逊 AWS 市场部郭多娇女士、何菁女士在写作和出版过程中提供的大力协助。

王毅 前亚马逊 AWS 解决方案架构师、区域主管



目 录

第1章 AWS 概览1

1.1 云计算是什么1

1.2 AWS 是什么2

1.3 AWS 有什么好处2

1.3.1 按需分配, 按用量付费2

1.3.2 弹性容量2

1.3.3 增加速度和灵活性2

1.3.4 全球性覆盖3

1.4 AWS 的安全措施3

1.5 服务概览4

1.5.1 全球基础架构4

1.5.2 服务概览5

1.6 怎样开始使用 AWS6

1.7 如何与 AWS 服务交互6

1.7.1 AWS 管理控制台6

1.7.2 AWS Command Line

Interface(AWS CLI)6

1.7.3 AWS 软件开发工具包(SDK)7

1.7.4 API 接口7

第2章 计算服务介绍9

2.1 什么是 EC2 弹性计算9

2.2 EC2 有哪些特点9

2.3 如何上手创建 EC2 实例10

2.3.1 选择亚马逊系统映像(AMI)10

2.3.2 实例类型11

2.3.3 安全组14

2.4 存储选项14

2.4.1 EBS14

2.4.2 实例存储14

2.5 实例的用户数据和元数据15

2.6 密钥对18

2.7 实例的状态19

2.7.1 实例启动19

2.7.2 停止和启动实例19

2.7.3 实例重启20

2.7.4 实例终止20

2.8 弹性负载均衡器(Elastic Load Balancer)21

2.8.1 什么是弹性负载均衡器21

2.8.2 ELB 能实现哪些功能21

2.9 自动伸缩组23

2.9.1 什么是自动伸缩组(Auto Scaling Group)23

2.9.2 ASG 的使用场景23

2.9.3 如何创建自动伸缩组24

第3章 Amazon RDS29

3.1 Amazon RDS 简介29

3.2 使用 RDS 可以带来的好处30

3.3 创建 Amazon RDS 数据库31

3.4	管理 Amazon RDS 数据库的存储	36
3.4.1	RDS 的存储类型	36
3.4.2	性能指标	36
3.4.3	SSD 存储	37
3.4.4	预配置 IOPS 存储	38
3.4.5	数据库实例类	40
3.5	Amazon RDS 数据库的备份与恢复	41
3.5.1	自动备份	41
3.5.2	使用数据库快照	42
3.5.3	将数据库恢复至某个时间点	46
3.6	Amazon RDS 数据库的安全性	49
第 4 章	Amazon VPC	51
4.1	VPC 介绍	51
4.2	Amazon VPC 的主要功能	52
4.3	VPC 的基本概念	53
4.4	Amazon VPC 基本操作	53
4.4.1	VPC 的建立及大小调整	53
4.4.2	删除 VPC	55
4.4.3	子网的建立	56
4.4.4	在子网中启动实例	58
4.4.5	删除子网	61
4.5	在 Amazon VPC 中设置路由表	61
4.5.1	路由表	61
4.5.2	路由表的基本信息	61
4.5.3	主路由表	62
4.5.4	自定义路由表	63

4.6	安全性与防火墙	63
4.6.1	VPC 基本安全	63
4.6.2	安全组与网络 ACL 的比较	64
第 5 章	Amazon CloudFront	65
5.1	什么是 Amazon CloudFront	65
5.2	Amazon CloudFront 服务的优势	66
5.2.1	快速	66
5.2.2	简便	66
5.3	创建 Web 分发	66
5.4	利用 CloudFront 提供私有对象的 CDN 服务	70
5.5	私有内容的工作原理	71
5.6	对 CloudFront 进行负载测试	72
第 6 章	Amazon DynamoDB	73
6.1	Amazon DynamoDB 介绍	73
6.2	使用 DynamoDB 能带来哪些好处	74
6.3	DynamoDB 入门	75
6.3.1	数据模型概念: 表、项目和属性	75
6.3.2	主键	76
6.3.3	二级索引	77
6.3.4	DynamoDB 数据类型	78
6.3.5	DynamoDB 支持的操作	78
6.3.6	预配置吞吐容量	80
6.3.7	访问 DynamoDB	80
6.4	创建 DynamoDB 表	81
6.4.1	准备工作	81
6.4.2	创建示例表	84

6.4.3 加载示例数据.....	88	8.3 Amazon EMR 使用案例.....	133
6.4.4 开始查询.....	97	8.3.1 点击流分析.....	133
6.5 DynamoDB 的最佳实践.....	98	8.3.2 基因学.....	133
6.5.1 表的最佳实践.....	98	8.3.3 日志处理.....	133
6.5.2 项目最佳实践.....	103	8.4 如何使用 Amazon EMR.....	133
6.5.3 查询和扫描最佳实践.....	104	8.5 创建 Amazon EMR 集群.....	134
6.5.4 Local Secondary Index 最佳实践.....	105	8.6 Amazon EMR 的概念.....	140
6.5.5 Global Secondary Index 最佳实践.....	107	8.6.1 Amazon EMR 有什么用途... 141	
第 7 章 AWS Elastic Beanstalk.....	109	8.6.2 Amazon EMR 的功能.....	142
7.1 什么是 AWS Elastic Beanstalk.....	109	8.6.3 Amazon EMR 是 如何工作的.....	145
7.2 为什么需要 Elastic Beanstalk.....	110	8.6.4 为 Amazon EMR 提供了什么 工具.....	149
7.3 Elastic Beanstalk 入门.....	110	8.7 Amazon EMR 上常用的 Hadoop 工具.....	150
7.3.1 Elastic Beanstalk 的组件.....	110	8.7.1 Hive 和 Amazon EMR.....	150
7.3.2 权限.....	114	8.7.2 Impala 和 Amazon EMR.....	151
7.3.3 支持的平台.....	120	8.7.3 Pig 和 Amazon EMR.....	153
7.3.4 设计注意事项.....	120	8.7.4 使用 HBase 存储数据.....	153
7.4 如何使用 Elastic Beanstalk.....	123	8.7.5 配置 Hue 以查看、查询或 操作数据.....	154
第 8 章 Amazon EMR.....	131	第 9 章 AWS CloudFormation.....	157
8.1 Amazon EMR 介绍.....	131	9.1 AWS CloudFormation 介绍.....	157
8.2 Amazon EMR 的优势.....	132	9.2 AWS CloudFormation 的 优势.....	158
8.2.1 易于使用.....	132	9.2.1 广泛支持 AWS 资源.....	158
8.2.2 成本低廉.....	132	9.2.2 易于使用.....	158
8.2.3 灵活.....	132	9.2.3 声明性和灵活性.....	158
8.2.4 运行可靠.....	132	9.2.4 基础设施即代码.....	158
8.2.5 安全.....	132	9.2.5 通过参数实现自定义.....	159
8.2.6 灵活.....	133	9.2.6 便于集成.....	159



9.3 什么是 AWS	
CloudFormation	159
9.3.1 AWS CloudFormation 的概念	160
9.3.2 AWS CloudFormation 是如何运行的	163
9.3.3 删除堆栈工作流程	165
9.4 使用堆栈	166
9.4.1 使用 AWS CloudFormation 控制台	166
9.4.2 使用 AWS 命令行界面	167
9.4.3 AWS CloudFormation 堆栈更新	167
9.4.4 与 Windows Stacks 共同运行	168
9.5 使用模板	169
9.5.1 模板剖析	169
9.5.2 示例模板	182
9.5.3 创建模板	186
9.6 最佳实践	197
9.6.1 按生命周期和所有权组织堆栈	198
9.6.2 使用 IAM 控制访问	199
9.6.3 验证所有资源类型的配额	199
9.6.4 重复使用模板以在多个环境中复制堆栈	199
9.6.5 使用嵌套堆栈来重复使用常见模板模式	199
9.6.6 请勿将证书嵌入模板	200
9.6.7 使用 AWS 特定的参数类型	200
9.6.8 使用参数约束	200
9.6.9 使用 AWS::CloudFormation::Init 在 Amazon EC2 实例上部署软件应用程序	200
9.6.10 在使用模板前验证模板	201
9.6.11 通过 AWS CloudFormation 管理所有堆栈资源	201
9.6.12 使用堆栈策略	201
9.6.13 使用 AWS CloudTrail 记录 AWS CloudFormation 调用	201
9.6.14 使用代码审查和修订控制来管理模板	202
9.7 使用 IAM 访问控制	202
9.7.1 AWS CloudFormation 操作和资源	202
9.7.2 AWS CloudFormation 控制台特定的权限	204
9.7.3 AWS CloudFormation 条件	205
9.7.4 确认 AWS CloudFormation 模板中的 IAM 资源	205
9.7.5 管理 Amazon EC2 实例上运行的应用程序的证书	206
9.7.6 授予临时访问权限(联合访问)	206
第 10 章 Amazon Kinesis	209
10.1 Amazon Kinesis 介绍	209
10.2 Amazon Kinesis 的优势	210
10.3 Amazon Kinesis 的使用场景	211
10.4 Amazon Kinesis 的概念	212
10.4.1 主要概念	212



10.4.2 Streams 技术.....	214	10.6 使用 Amazon Kinesis	226
10.4.3 创建器	215	10.6.1 将数据写入流	226
10.4.4 使用器	216	10.6.2 读取流中的数据	230
10.5 Amazon Kinesis 入门.....	217	10.6.3 监控	235
10.5.1 设置	217	10.6.4 为流添加标签	235
10.5.2 基本操作	218	10.6.5 控制访问权限	236

第1章 AWS 概览

1.1 云计算是什么

云计算的定义,是指通过互联网向企业或家庭用户提供计算资源、存储资源、网络资源、应用资源等。云计算是一种分布式计算模式,它通过互联网将大量的计算资源集中起来,为用户提供按需使用的服务。云计算的特点包括:弹性伸缩、按需付费、自助服务、高可用性、安全性和隐私性等。云计算的应用场景非常广泛,包括企业级应用、个人应用、物联网、大数据、人工智能等。云计算已经成为当今信息技术领域的重要发展方向。

云计算是一种通过网络方式提供可扩展的计算和存储资源,包括:服务器、网络、存储、应用等。云计算的特点包括:弹性伸缩、按需付费、自助服务、高可用性、安全性和隐私性等。云计算的应用场景非常广泛,包括企业级应用、个人应用、物联网、大数据、人工智能等。云计算已经成为当今信息技术领域的重要发展方向。