

系统论述云存储安全的基础知识与关键技术  
深入解读大数据的必备知识  
业内知名专家联袂推荐



开发者书库



Cloud Storage Security  
Foundations of Big Data Analysis and Computing

# 云存储安全

## 大数据分析计算的基石

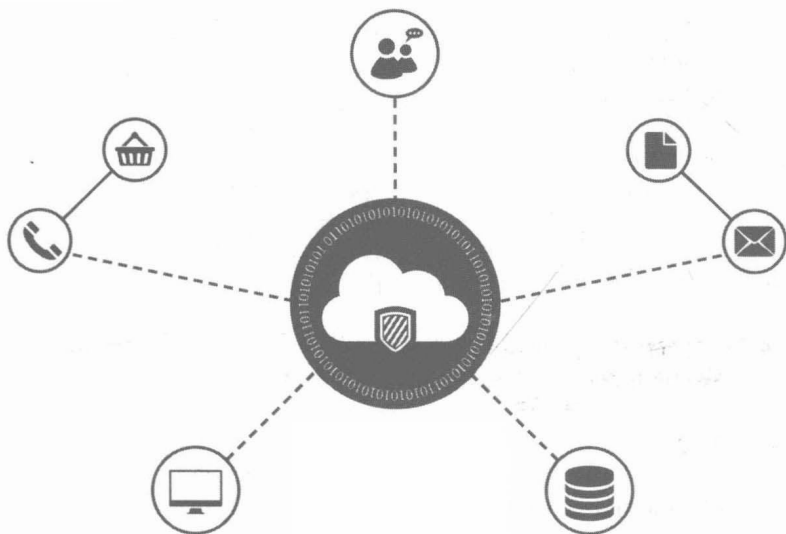
陈兰香◎著  
Chen Lanxiang

清华大学出版社



清華

开发者书库



Cloud Storage Security  
Foundations of Big Data Analysis and Computing

# 云存储安全

大数据分析计算的基石



陈兰香◎著  
Chen Lanxiang

清华大学出版社  
北京

## 内 容 简 介

本书系统而全面地介绍了云存储安全相关的关键技术及其最新研究成果。首先对云存储做一概述；然后从云存储安全体系结构说起，按照云存储安全的需求层次，依次介绍云存储虚拟化安全、云存储系统身份认证与访问控制、加密云存储系统、密文云存储信息检索、云存储服务的数据完整性审计、云存储数据备份与恢复等内容；最后介绍大数据时代的云存储安全。

云存储服务是大数据时代数据存储的基础，保障云存储安全是大数据分析计算的基石。本书内容由浅入深，按照云存储安全的需求层次以及保障数据安全的逻辑层次，对关键技术逐一进行介绍。全书共分9章，每章都是从概述开始，根据需求逐步介绍，主要是最前沿的成果，然后对相关领域的研究工作进行总结，指出存在的问题及将来的研究方向。

本书作者长期从事云存储安全的相关研究工作，对该领域的前沿科研成果比较熟悉。本书内容极具参考价值，对于信息安全相关专业的本科生及研究生具有很好的指导意义，可以帮助他们全面系统地学习云存储安全领域的基础知识和前沿成果，建立保障大数据安全的存储体系。

本书可作为高等院校信息安全、网络空间安全、信息存储、计算机科学与技术、密码学与信息对抗等相关专业的本科生和研究生教材，也可作为通信工程师和计算机网络工程师的参考读物，对于从事信息安全领域研究工作的科研人员也有很好的指导意义和参考价值。

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

云存储安全：大数据分析计算的基石/陈兰香著. —北京：清华大学出版社，2019

(清华开发者书库)

ISBN 978-7-302-53119-7

I. ①云… II. ①陈… III. ①计算机网络—信息存储—信息安全—研究 IV. ①TP393.071

中国版本图书馆 CIP 数据核字(2019)第 100740 号

责任编辑：盛东亮 钟志芳

封面设计：李召霞

责任校对：梁毅

责任印制：丛怀宇

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](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈：010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载：<http://www.tup.com.cn>, 010-62795954

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：186mm×240mm 印 张：15.75

字 数：364千字

版 次：2019年9月第1版

印 次：2019年9月第1次印刷

定 价：69.00元

产品编号：081387-01

# 前言

## PREFACE

---

图灵奖获得者吉姆·格雷(Jim Gray)在其获奖演说中指出:由于互联网的发展,未来18个月新产生的数据量将是有史以来数据量之和。从预言至今,数据量的增长基本符合这个定律。

人类社会产生的数据信息一方面来自于互联网;一方面来自于日常生产、生活及各种科学试验,例如科学计算和仿真、飞行动力学、核爆炸仿真、太空探测及医学影像等每天所产生的数据量更是大到惊人的程度。

根据易观智库发布的《中国大数据市场年度综合报告2016》中数据显示,2015年中国大数据市场规模达到105.5亿元,同比增长39.4%,预计未来3~4年,市场规模增长率将保持在30%以上。

云存储作为大数据时代的存储基础设施,其重要性不言而喻,特别是物联网技术的高速发展,其后的支撑平台也有赖于云存储技术。在已经实现的云存储服务中,数据安全和隐私保护问题一直令人担忧,并已经成为阻碍云存储发展和推广的主要因素之一。从现实情况看,云存储数据安全问题层出不穷。

2014年9月,黑客利用苹果 iCloud 云端系统的漏洞将其数据外泄;2015年4月,上海、重庆等超30个省市约5000万用户社保信息被泄露;2016年4月,土耳其方面爆发重大数据泄露事件,导致近5000万土耳其公民的个人信息遭到威胁;2017年2月,知名云安全服务商 Cloudflare 被曝泄露用户 HTTPS 网络会话中的加密数据长达数月;2018年1月,印度10亿公民身份数据库 Aadhaar 被曝遭到网络攻击,除了名字、电话号码、邮箱地址等信息之外,指纹、虹膜记录等极度敏感的信息均遭到泄露……各类安全事故不胜枚举。

为了推进云存储技术的快速发展与普及,本书全面、系统地介绍了云存储安全的发展历程和最新研究成果。

在信息安全的三要素(CIA三元组)——机密性(Confidentiality)、完整性(Integrity)、可用性(Availability)的基础上,作者认为应加入访问控制(Access Control),将CIA延伸到CIAA,此四方面被认为是保障云存储安全的核心技术。因此,本书将围绕此四方面及其衍生的其他问题展开讨论,全书共分为9章。第1章对云存储进行概述,介绍云存储的兴起与存储安全面临的挑战;第2章建立云存储安全体系结构,围绕云存储系统安全体系结构说明本书的研究内容;第3章介绍云存储虚拟化安全;第4章介绍云存储系统身份认证与访问控制;第5章介绍加密云存储系统;第6章介绍密文云存储信息检索;第7章介绍云存

储服务的数据完整性审计；第8章介绍云存储数据备份与恢复；第9章详细阐述大数据时代的云存储安全。

本书主要针对已有一定信息安全相关基础知识的读者，比如知道密码技术，能区分对称密码与公钥密码，知道当前使用的对称密码标准是什么以及常用的公钥密码技术；知道Hash算法、消息认证码(Message Authentication Code, MAC)等相关基础知识。关于密码技术的书籍和资料非常丰富，本书没有再介绍相关理论知识。

本书取材新颖，结构合理，不仅包括云存储安全技术的基础理论，而且涵盖了云存储安全技术的最新研究成果，力求使读者通过本书的学习了解本学科最新的发展方向。本书适合作为高等院校信息安全、网络空间安全、信息存储、计算机科学与技术、密码学与信息对抗等相关专业的本科生和研究生教材，也可作为通信工程师和计算机网络工程师的参考读物。

因为本书内容涉猎广泛，所以难免存在一些疏漏或考虑不周全、引用不全之处，但作者绝对是本着讲授本领域最新研究成果的想法，尽可能地介绍本书各部分内容的精华或卓越观点，通过通俗易懂、深入浅出的讲解，既可以实现传播知识的科普目标，也可将其作为“引子”为入门者抛砖引玉，以实现登堂入室之目的。因本人知识见闻有限，难免有“趋熟避生”之嫌，又或者“词不达意”“言不尽意”，让读者产生误解。希望读者能够谅解，并在方便之时让我知晓，使我有机会给予解释，同时交流学习，以待以后有机会更正。

非常希望此书能够做到开卷有益！

作者

2019年5月

# 目录

## CONTENTS

---

<b>第 1 章 云存储概述</b> .....	1
1.1 云存储的兴起 .....	1
1.2 云存储发展现状 .....	3
1.2.1 定义、服务模型与分类 .....	3
1.2.2 为什么需要云存储.....	5
1.2.3 现状与发展趋势.....	8
1.3 云存储安全.....	11
1.3.1 为什么有安全问题 .....	11
1.3.2 云存储安全威胁 .....	12
1.3.3 需要解决的几个问题 .....	16
1.3.4 面临的挑战 .....	17
1.4 本章小结.....	18
参考文献 .....	18
<b>第 2 章 云存储安全体系结构</b> .....	20
2.1 云存储安全体系.....	20
2.1.1 云存储系统层次模型 .....	20
2.1.2 云存储系统安全体系结构 .....	21
2.2 数据生命周期中的安全风险.....	23
2.3 保障云存储安全的几个原则.....	25
2.4 本章小结.....	26
参考文献 .....	26
<b>第 3 章 云存储虚拟化安全</b> .....	28
3.1 云存储虚拟化技术.....	28
3.1.1 虚拟化技术概述 .....	29

3.1.2 虚拟化技术分类 .....	31
3.1.3 虚拟化带来的安全挑战 .....	34
3.2 针对虚拟化的安全攻击 .....	34
3.2.1 虚拟机攻击分类 .....	35
3.2.2 虚拟机攻击方法 .....	36
3.3 虚拟机安全机制 .....	41
3.3.1 虚拟机访问控制 .....	41
3.3.2 虚拟机隔离 .....	43
3.3.3 其他安全机制 .....	46
3.4 存在的问题与未来发展方向 .....	48
3.5 本章小结 .....	49
参考文献 .....	49
<b>第4章 云存储系统身份认证与访问控制 .....</b>	<b>57</b>
4.1 身份认证与访问控制概述 .....	57
4.1.1 基础知识 .....	57
4.1.2 传统访问控制 .....	58
4.1.3 云存储系统的访问控制 .....	61
4.2 相关理论知识 .....	62
4.2.1 双线性对 .....	62
4.2.2 访问结构 .....	63
4.2.3 基于属性加密机制 .....	64
4.3 云存储系统访问控制相关研究 .....	67
4.3.1 研究概述 .....	67
4.3.2 基于对称密码的访问控制 .....	68
4.3.3 基于属性加密的访问控制 .....	69
4.3.4 产业界的实践 .....	73
4.3.5 其他 .....	75
4.4 存在的问题与未来发展方向 .....	76
4.5 本章小结 .....	77
参考文献 .....	77
<b>第5章 加密云存储系统 .....</b>	<b>84</b>
5.1 云环境下加密存储面临的挑战 .....	85

5.2	加密云存储系统的发展	86
5.2.1	加密存储系统发展历程	87
5.2.2	产业界的实践	91
5.3	数据共享中密钥管理	93
5.3.1	密钥生成与发布	93
5.3.2	密钥撤销	95
5.4	密文重复数据删除	97
5.5	加密云数据库	102
5.6	存在的问题与未来发展方向	103
5.7	本章小结	104
	参考文献	104
<b>第6章</b>	<b>密文云存储信息检索</b>	<b>110</b>
6.1	密文搜索技术概述	110
6.1.1	密文搜索技术分类	110
6.1.2	密文搜索应用模型	112
6.2	密文搜索发展现状	115
6.2.1	密文搜索功能属性与安全属性	115
6.2.2	密文搜索现状与发展趋势	116
6.3	云存储环境下密文搜索	119
6.3.1	云存储环境下的特殊需求	119
6.3.2	最新云存储密文搜索方案	120
6.4	未来发展方向	133
6.5	本章小结	134
	参考文献	134
<b>第7章</b>	<b>云存储服务的数据完整性审计</b>	<b>140</b>
7.1	数据完整性审计概述	140
7.1.1	问题的起源	140
7.1.2	完整性审计方案分类	141
7.1.3	完整性审计目标	144
7.2	云存储中数据完整性审计发展现状	146
7.2.1	完整性审计框架	146
7.2.2	云存储环境下的需求	147



7.2.3 发展现状与趋势·····	147
7.3 最新完整性审计方案·····	157
7.4 未来发展方向·····	159
7.5 本章小结·····	160
参考文献·····	161
<b>第8章 云存储数据备份与恢复·····</b>	<b>167</b>
8.1 数据备份与恢复概述·····	167
8.1.1 备份系统分类·····	167
8.1.2 性能指标·····	168
8.2 纠删码技术·····	169
8.2.1 纠删码原理·····	170
8.2.2 纠删码的发展·····	170
8.3 数据备份技术·····	173
8.3.1 基于多副本的备份·····	173
8.3.2 基于纠删码的备份·····	177
8.3.3 几种备份技术的优缺点·····	179
8.4 数据恢复技术·····	182
8.5 其他相关研究·····	183
8.6 举例：基于喷泉码的数据备份与恢复·····	185
8.6.1 基于喷泉码的编码方法·····	185
8.6.2 错误检测方法·····	187
8.6.3 数据恢复方法·····	191
8.7 本章小结·····	192
参考文献·····	193
<b>第9章 大数据时代的云存储安全·····</b>	<b>199</b>
9.1 大数据概述·····	200
9.1.1 基本概念·····	200
9.1.2 大数据带来的数据存储挑战·····	201
9.1.3 大数据的应用价值·····	203
9.2 大数据环境下的云存储安全·····	205
9.2.1 安全挑战·····	205
9.2.2 密态计算·····	207

9.2.3 安全多方计算·····	210
9.2.4 隐私保护·····	214
9.2.5 举例：健康医疗大数据安全保护·····	218
9.3 基于 NoSQL 的大数据云存储·····	222
9.4 基于区块链的大数据云存储·····	225
9.4.1 区块链概述·····	225
9.4.2 基于区块链技术保障大数据安全·····	226
9.5 存在的问题和未来发展方向·····	229
9.6 本章小结·····	232
参考文献·····	232



# 云存储概述

《三国演义》第一回：“话说天下大势，分久必合，合久必分……”。数据存储系统也不例外，从传统分散式存储系统，发展到集中式存储，然后发展到现在的集中式云存储，又往分布式云存储系统方向发展……

云存储(Cloud Storage)是在云计算(Cloud Computing)概念上延伸和发展起来的，是指通过集群应用、网格技术和分布式文件系统等功能，将网络中大量不同类型的存储设备通过虚拟化软件集合起来协同工作，实现共同对外提供数据存储和业务访问功能。当云计算系统处理的核心是大量数据的存储和管理时，云计算系统就需要配置大量的存储设备，那么云计算系统就转变成云存储系统，所以云存储是一个以数据存储和管理为核心的云计算系统。

本章将从云存储的兴起讲起，详细介绍云存储的发展现状与趋势，然后详细说明为什么会存在安全问题，具体有哪些安全威胁，又有哪些需要解决的问题，解决了这些问题仍然面临怎样的挑战。

本章是为后续章节作一个铺垫。通过本章的介绍，用户可以了解到云存储安全技术要解决的问题；具体研究范围，后续章节将一一展开讨论。

## 1.1 云存储的兴起

云存储的兴起可以从一个趣闻说起。全球最大网上书店亚马逊( Amazon, [www.amazon.com](http://www.amazon.com))是一个电子商务平台，早期的网络服务平台 Obidos 采用 C++ 语言编写，编译后的代码大小为 700MB，编译一次需要一天时间，使加入新功能变得越来越困难。后来，他们设计并实现了一个新的服务平台 Gurupa，采用基于 Perl 语言的 Mason 模板库，把所有功能以微服务的形式集成起来，但是性能不好。为了应对圣诞节的流量高峰期，亚马逊购买了大量服务器和 Cisco 交换机，用以实现负载均衡，以满足流量高峰时对性能扩展的需求。但是，节日过后的淡季，又不得不面临大量机器空闲的状况。为了不让资源闲置，亚马逊就把这些机器配置成服务来租赁，这就是最初的云计算的雏形。

利用已有的 IT 基础设施——硬件设备、服务器与交换机，组合配置成集计算、存储与

网络于一体的资源池,一方面可为电子商务平台提供各类 IT 服务,满足各种负载的需求;另一方面还可将闲置的资源分解成一个个小单元用于租售,实现成本的分摊。正是看到了这一点,亚马逊利用虚拟化技术——云计算与云存储的核心技术,将闲置的 IT 资源进行分解,在其上构建了亚马逊网络服务系统(Amazon Web Services, AWS)。2002 年 7 月,亚马逊利用其分布在全球各地的数据中心,推出面向第三方的云计算服务 AWS,主要包括数据库服务、处理器资源租赁、网络存储、应用软件服务等。AWS 的迅速成长让其成为亚马逊的一项非常成功的新业务。

亚马逊的创始人 Jeff Bezos 在一次采访中说过:亚马逊作为电子商务公司,起初为了处理大量的货品库存和分配,积累并完善了他们的大数据计算技术。目前,亚马逊提供的服务包括:亚马逊弹性计算云(Amazon Elastic Compute Cloud, EC2)、亚马逊简单存储服务(Amazon Simple Storage Service, S3)、亚马逊 Web 服务(Amazon Web Services)、亚马逊简单数据库(Amazon SimpleDB)、亚马逊简单队列服务(Amazon Simple Queue Service)以及亚马逊内容分发网络(Amazon CloudFront)等。

回顾历史,任何事物的发展都存在一定的偶然性和必然性。在 Brad Stone 于 2013 年撰写的关于亚马逊历史最权威的 *The Everything Store: Jeff Bezos and the Age of Amazon*<sup>[1]</sup> 一书中,可以归纳出影响亚马逊发展的历史必然性的几个因素<sup>[2]</sup>。

亚马逊的核心业务——电子商务有很强的季节性。2002—2003 年,公司发展进入了瓶颈期,如何有效配置兼顾扩展性与持续性的基础服务平台成为一个亟待解决的问题。而这个问题,在当时只有亚马逊才存在,其他公司如谷歌(Google)当时的营利模式主要是投放广告业务,所以亚马逊具备开发云计算服务所需要的发展动力和生存压力。

2002 年, Tim O'Reilly(O'Reilly Media 出版公司的创始人)拜访 Bezos, 希望与亚马逊合作。合作没谈成,但 O'Reilly 的提议让 Bezos 意识到亚马逊的数据可以开放给第三方程序员使用,于是他组织了第一届亚马逊开发者大会,提出所有互操作要以 API(Application Programming Interface, 应用编程接口)的方式提供数据和各种功能,而且 API 可以对外部人员开放, AWS 就是 API 化的服务平台,这种方式为后面的系统扩展性打下了良好的基础。

Bezos 当时对图书(*Creation: Life and How to Make It*)<sup>[3]</sup> (2001 年出版)非常着迷,并且让公司高管人手一册。此书作者 Steve Grand 无意中给亚马逊高管指出了一条解决 IT 资源配置的思路:把 IT 基础设施分成一个个小单元,让程序员可以自由配置与使用。因此, Bezos 马上组建研发团队来研究开发这样的小单元,这便是虚拟化思想的启蒙。

2004 年,亚马逊负责 IT 基础设施配置的 Chris Pinkham 希望回到老家南非。为了挽留他,亚马逊就在好望角设置了一个办公室,让 Pinkham 可以远程办公。为了能跟西雅图的总部一直保持连线, Pinkham 带领一个程序员 Chris Brown 开发了最早版本的 EC2 和 S3。

2006 年,亚马逊的董事会和硅谷风投并不看好 AWS,因为这看起来跟电子商务的主营业务完全没有关系,而且那时电子商务的主要产品——书籍一直在亏钱。按照董事会正常

的商业逻辑,会否决上线 AWS 这样疯狂而大胆的新产品,但是在亚马逊的董事会上 Bezos 拥有极强的影响力;而且当时他已经成功地运营了极具前瞻性的众包产品土耳其机器人 (Amazon Mechanical Turks),向董事会证明亚马逊可以走出主营业务,开发出成功的新产品。这些因素使得 Bezos 可以说服董事会,继续发展 AWS。

最后也是最重要的一个因素是,Bezos 一直坚信亚马逊的价值在于提供近似于水电的基础设施服务,这样才可以更好地服务于用户。他认为成本应该越低越好,这与 Steve Jobs 的商业逻辑刚好相反:Jobs 是高价+小量+超额利润,而 Bezos 是超低价+巨量+微薄利润。在这样的思想指导下,AWS 一开始的价格非常低,因为 Bezos 没准备在短期内盈利,而且他刻意压低价格,不想引起潜在竞争对手的注意。而 Google 的主营业务——广告的利润非常高,在当时既没有压力也没有动力,更加没有说得过去的理由来介入一个看起来根本不赚钱的生意。Google 的 Eric Schmidt 说:他在两年里,发现很多新兴公司都在用同一家公司——亚马逊的服务,这才让他意识到亚马逊已经在下一盘很大的棋。

所有的科技进步都是在解决现实问题的同时提供更好的用户体验,亚马逊做到了,Bezos 非常注重客户体验。同时,亚马逊本身就有全球化的数据中心,这与是否存在 AWS 没有关系,但 AWS 服务将更多的客户带入亚马逊,因此 AWS 与传统的电子商务系统协同发展,相互促进,这也是亚马逊云计算技术得以突飞猛进的一个重要原因。

此后,微软的 Azure 和谷歌的应用引擎(App Engine)都在尝试亚马逊的这种商业模式。中国的百度云、阿里云等也赶上了这次云计算浪潮,目前也初具规模。

目前,信息存储系统还朝着无限的带宽、无限的容量和无限的处理能力 (Infinite Bandwidth, Infinite Capacity, Infinite Processing Capability),即 3I 的方向飞速发展,其目标是实现“Anytime, Anywhere, Anything”3A 目标,即可在任意时间、任意地点实现任意数据访问。存储产品不再是附属于服务器的辅助设备,而成为互联网中最主要的花费所在。信息技术正从以计算为核心的计算时代进入到以存储为核心的存储时代,网络化存储已经成为存储市场的热点。而目前的云存储服务是网络存储发展的必然趋势。

## 1.2 云存储发展现状

云存储是一个以数据存储和管理为核心的云计算系统,云存储与云计算息息相关。

### 1.2.1 定义、服务模型与分类

2011年9月,美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)<sup>[4]</sup>对云计算的定义、特征、服务模式和类型作了详细说明。

云计算是一种商业计算模型,它可以实现随时随地及按需应变的可配置的 IT 资源(例如,计算、存储、网络、服务器、应用),资源能够快速供应并释放,使管理资源的工作量及与服务提供商的交互减小到最低限度。它将计算任务分布在大量计算机构成的资源池上,使各种应用系统能够根据需要获取计算力、存储空间和各种软件服务。它是并行计算(Parallel

Computing)、分布式计算(Distributed Computing)和网格计算(Grid Computing)的发展,或者说是这些计算机科学概念的商业实现。

有计算的地方就有存储,特别是在大数据时代,数据为王,通常需要将计算能力迁移到存储端,比如最近提出的 Near-Data Processing(近数据端处理)、In-Data Processing(在数据端处理)、Processing-in-Memory(在内存中处理)及 Processing-in-Storage(在存储中处理),存储与计算越来越不可分离。因为数据量太大,将数据迁移到计算端的时间可能比直接将存储数据的设备使用卡车运送到计算端还要慢。

云计算的服务模型可以分为 3 种,如图 1-1 所示。

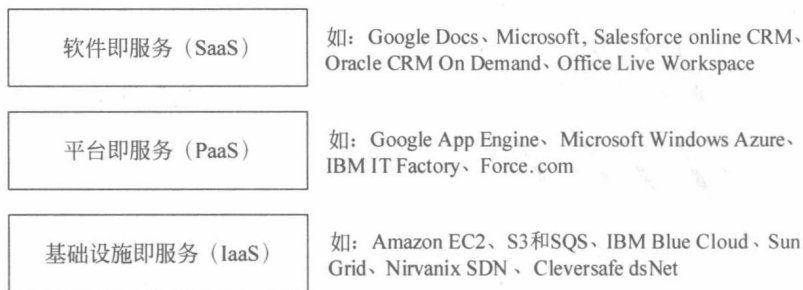


图 1-1 云计算的 3 种服务模型

- 软件即服务(Software as a Service, SaaS): 是一种通过互联网提供软件的模式,用户无须购买软件,可直接使用构建在云端的软件来管理企业经营活动。在这一方面,比较典型的有 Google Docs、Microsoft、Salesforce online CRM、Oracle CRM On Demand、Office Live Workspace 等。
- 平台即服务(Platform as a Service, PaaS): 用户使用云平台所支持的语言和工具,开发应用并部署在云平台上。用户不直接管理或控制包括网络、服务器、存储,甚至单个应用功能在内的底层云基础设施,但可以控制部署的应用程序,也有可能配置应用的托管环境。比如将软件开发平台作为一种服务,以 SaaS 的模式提交给用户。PaaS 的出现可以加快 SaaS 的发展,尤其是加快 SaaS 应用的开发速度。在这一方面,比较典型的有 Google App Engine、Microsoft Windows Azure、IBM IT Factory、Force.com 等。
- 基础设施即服务(Infrastructure as a Service, IaaS): 用户通过互联网可以获得完善的计算机基础设施服务。5G 是高效、高速的移动互联的基础设施,随着未来 5G 技术的发展,对基础设施服务的需求会日益增长。比如提供处理器、存储、网络等(虚拟)硬件资源给用户,用户可任意安装软件和开发环境,包括安装操作系统和应用程序。用户不管理或控制底层的基础设施,但可以控制操作系统、存储、部署的应用,也有可能选择网络构件(例如,主机防火墙)。在这一方面,比较典型的有亚马逊 EC2、S3 和 SQS、IBM Blue Cloud、Sun Grid、Nirvanix SDN、Cleversafe dsNet 等。

这3种模型从应用到平台再到架构,越来越底层,开发者获得的可操作性和灵活性也越来越大。通常说的云存储一般可分类到IaaS,但对于云存储服务提供者,他们提供的PaaS和SaaS同样需要云存储技术来部署相应的平台。

按照部署方式,云计算可以分为私有云(Private Cloud)、社区云(Community Cloud)、公共云(Public Cloud)与混合云(Hybrid Cloud)4种模式。

私有云是指构建在一个组织内部且为该组织或者信任该组织的用户提供服务的云,可以由该机构或第三方管理;社区云是指一些有着共同利益(如任务、安全需求、策略、规约考虑等)并打算共享基础设施的组织共同创立的云,可以由该机构或第三方管理;公共云是指若干企业和用户共享使用的一种云环境,由销售云服务的组织机构管理;混合云由两个或两个以上的云(私有云、社区云或公共云)组成,它们各自独立,但通过标准化技术或专有技术绑定在一起,云之间实现了数据和应用程序的可移植性。

云计算与云存储密不可分,因此云计算的定义、服务模型和分类同样适用于云存储。下面将介绍为什么需要云存储。

## 1.2.2 为什么需要云存储

据国际数据公司(International Data Corporation, IDC)2013年的报告<sup>[5]</sup>显示,2012年全球数据已经达到2.8ZB(1ZB等于1万亿GB,2.8ZB也就相当于28亿个1TB的移动硬盘),而这个数值还在以每两年翻一番的速度增长,预计到2020年全球将总共拥有40ZB的数据量,如图1-2所示。

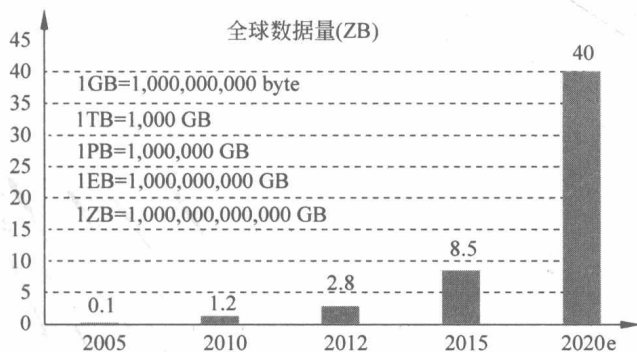


图 1-2 IDC 预测全球数据量增长趋势

而2013年中国的数据量占比为13%,数据总量超过0.8ZB(相当于8亿TB),2倍于2012年,相当于2009年全球的数据总量。预计到2020年,中国产生的数据总量将是2013年的10倍,超过8.5ZB。2013全球数据分布如图1-3所示。

全球IT市场咨询公司Springboard Research于2010年6月10日发布了《中国云存储服务报告》(China Cloud Storage Services Report)<sup>[6]</sup>。报告显示,未来5年中国云存储服务市场的年复合增长率将达到103%,平均每年市场价值翻一番。从图1-4中可以看出,中国云存储服务的市场价值将由2009年的605万美元快速增长至2014年的2.0854亿美元。

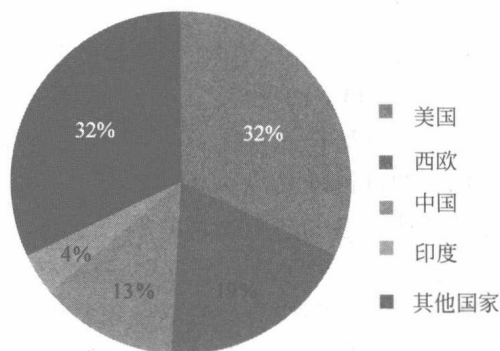


图 1-3 2013 全球数据分布

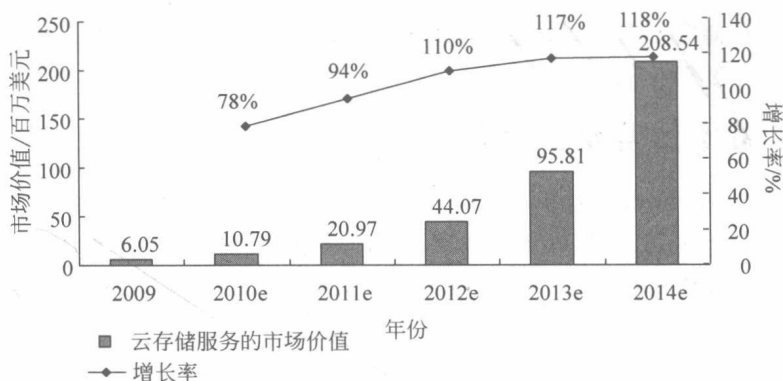


图 1-4 2009—2014 年中国云存储服务的市场价值

同时报告指出,尽管每月每 GB 的存储服务价格持续下降,但是云存储市场总容量的增长幅度更快,从而推动云存储市场整体规模在未来 5 年内的快速上涨。图 1-5 显示了 2009—2014 年中国云存储服务的存储容量需求,预计将从 2009 年的 0.6PB 上升到 2014 年的 66.29PB,增长了 110 倍以上。

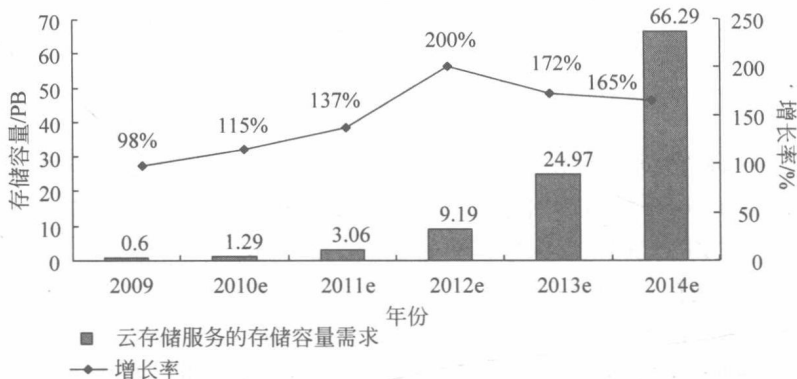


图 1-5 2009—2014 年中国云存储服务的存储容量需求



因为最新数据未公开,所以图示数据都是旧的数据,在本小节中只是以此说明数据量在呈指数级增长,中国在全球数据量的占比是比较高的,云存储市场潜力巨大。

根据 IBM 的调查统计报告<sup>[7]</sup>,企业的 IT 费用呈逐年上升趋势,如图 1-6 所示。该调查报告将 IT 费用分解为 3 个方面:新购置服务器的费用、服务器管理和维护费用、能源以及制冷设备的费用。在这 3 个方面中,服务器管理和维护费用开销最大,而且上升速度最快。为了保证业务高峰时 IT 系统的稳定性,企业实际部署的服务器的峰值工作量比平均值要高 2~10 倍,因此数据中心服务器的利用率一般只有 5%~20%。另外,在进行 IT 建设时,IT 工作人员花费 70% 的时间和精力做基础架构、软件以及日常的维护工作,只有 20% 或者更少的时间花在真正与业务相关的系统建设上。

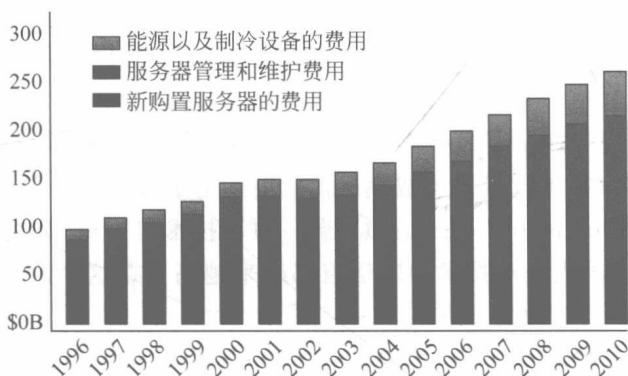


图 1-6 企业每年的 IT 费用统计与预测

虽然购置服务器和能源及制冷设备的成本相当,但是企业的管理和人员成本太高,利用率又太低。

选择云存储服务,一方面能够为企业的数据中心节省成本,还能够实现资源的集中共享,把空闲时段的资源补充到企业更需要的应用上去,也免去了日常的管理与维护费用,与其适配的能源及制冷设备亦可免去。源自云存储服务的规模经济性,可以实现更低的硬件成本、更低廉的电力价格、更低的管理费用,加上更高的利用率,使云存储服务的经济性提高达 30 倍<sup>[8]</sup>。

图灵奖获得者(Jim Gray)在其获奖演说<sup>[9]</sup>中指出:由于互联网的发展,未来每 18 个月新产生的数据量将是有史以来数据量之和。这说明人们对存储容量的需求是惊人的,存储市场具有无限的潜力。云存储是信息存储的一种趋势,它可为用户带来如下好处。

(1) 无须购置初始耗资较大的服务器,也免去了专业的服务器及数据管理人员,避免过大的初始投资,能源及制冷设备减少。

(2) 实现任意地点、任意时间、任意数据访问。

(3) 提供可用性、可维护性与扩展性保障。

(4) 保障法规遵从的需求。

(5) 实现数据长期保存。