

大数据创新人才培养系列

云计算与大数据

THE ESSENTIAL CRITERIA
OF CLOUD COMPUTING AND BIG DATA

◎ 孙宇熙 编著

云计算改变 IT，大数据改变商业

不是公有云才是核心技术，大数据也不仅仅是 Hadoop

本书追根溯源，揭开云计算与大数据的神秘面纱

 中国工信出版集团

 人民邮电出版社
POSTS & TELECOM PRESS

系列

云计算与大数据 (CITP) 系列

主编人：孙宇熙 著者：孙宇熙 孙宇熙 孙宇熙
ISBN 978-7-111-43212-0
2016年12月第1版

① 云计算 ② 大数据 ③ 云计算与大数据
④ 云计算与大数据 ⑤ 云计算与大数据

中国版本图书馆CIP数据核字(2016)第217759号

内容简介

本书以云计算和大数据为研究对象，从云计算和大数据的定义、特点、应用等方面进行了系统的阐述。本书共分5章，第1章介绍云计算和大数据的基本概念；第2章介绍云计算的架构和关键技术；第3章介绍大数据的存储和计算；第4章介绍云计算和大数据在商业中的应用；第5章介绍云计算和大数据在政府和社会中的应用。本书可作为高等院校计算机专业及相关专业的教材，也可供从事云计算和大数据工作的工程技术人员参考。

云计算与大数据

THE ESSENTIAL CRITERIA OF
CLOUD COMPUTING AND BIG DATA

◎ 孙宇熙 编著

人民邮电出版社
北京

图书在版编目 (C I P) 数据

云计算与大数据 / 孙宇熙编著. — 北京 : 人民邮电出版社, 2016. 12
ISBN 978-7-115-43572-9

I. ①云… II. ①孙… III. ①云计算 IV.
①TP393.027

中国版本图书馆CIP数据核字(2016)第217169号

内 容 提 要

本书以真实的案例和数据为基础, 详细介绍了云计算与大数据的基础知识和主要技术。全书共 5 章, 内容包括揭秘云计算、揭秘大数据、云计算与大数据体系架构剖析、云计算与大数据进阶、大数据应用与云平台实战。

本书理论联系实际, 避免了深奥的理论推导, 语言尽可能通俗易懂, 是一本对读者开卷有益的关于云计算和大数据的普及读物。

本书既适合希望了解云计算和大数据知识的广大读者阅读, 也可供对云计算和大数据知识感兴趣的 IT 专业人士、相关项目的决策者与领导者阅读参考, 亦可作为高等院校相关专业的教材或教学参考用书。

-
- ◆ 编 著 孙宇熙
责任编辑 邹文波
责任印制 沈 蓉 彭志环
 - ◆ 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>
固安县铭成印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 13 2016 年 12 月第 1 版
字数: 320 千字 2016 年 12 月河北第 1 次印刷
-

定价: 49.80 元

读者服务热线: (010) 81055256 印装质量热线: (010) 81055316

反盗版热线: (010) 81055315

过去的十年是各行各业见证科技颠覆与快速迭代的十年，如果要三言两语说清楚到底是怎样的科技在本质上改变了人们的生活方式、工作方式和社交方式，大多数人也许会说是移动互联网与社交网络，没错，从最终交付到用户的途径来看的确如此。但是如果要从科技的本质上去探讨，我会说是云与大数据在颠覆这一切。云计算只是构成云生态系统的几个重要支柱之一，另外还有云存储、网络、云管理与协同（M&O）、云安全等，缺一不可。不过为了让读者可以一目了然地了解本书所涉猎的内容，笔者还是在书名及行文中保留了云计算的字样。

笔者每天在业界与各行各业的人打交道，时常遇到行业内的人对云计算与大数据有五花八门的观点、需求与问题，发现有些观点、看法与理解是被“误导”的，很多业务需求和对问题的理解与抽丝剥茧后的事实本质有较大偏差，久而久之就有了结合云与大数据两大议题写书的想法，以期达到解惑的目的。

至于本书内容，笔者先举几个例子。

(1) 大数据之深入人心是近两三年的事情，街头巷尾可谓尽人皆知，只要说是做大数据的，人家一定问你是做 Hadoop 的吗。于是乎，你要不是专攻 Hadoop 的，你都不好意思跟人家说你在做大数据。那么 Hadoop 能解决所有的大数据的问题吗？答案：当然不是。本书将详细说明为什么大数据不仅仅是 Hadoop。

(2) 云计算比大数据要早四五年出名，从个人到企业到政府全都蜂拥而来，市场上名头最响的就是那些公有云的服务提供商了，于是有一种普遍性的观点是，不做公有云的（比如私有云、混合云等）就没有掌握云计算的核心科技。事实并非如此！本书从行业与科技发展的来龙去脉，拿数据与事实说话来为大家讲述云层下的故事。

(3) 再如软件化与商业化硬件平台，市场上一种普遍的观点认为软件的能力与灵活性无限而硬件的价值创新已经无足轻重，于是所有的数据中心中全面铺设的是基于 X86 架构的商用硬件平台（Commodity Hardware & COTS, COTS 即 Commercially/Commodity Off-the-Shelf）。此种做法值得商榷，两个观点：软件的能力极限是受到下层硬件的限制的；商用硬件架构显然不能解决所有的业务问题，并且也不是最好的（效率最高、性价比最高）问题解决之道。在书中，笔者对 COTS 的叫法提出了一种不同的声音：VDH（Volume-Discounted Hardware），本质上这才是“互联网+”时代的商业硬件的最终形态——多买多折扣。

那么，随着云计算和大数据风起云涌，我们今天各行各业遇到的挑战与机遇到底是什么？是云计算或大数据系统体系架构的设计与实现，还是最终应用的设计与交付，或是以上两大问题之间各层平台化服务架构的整合与搭建？本书结合笔者在工作实践中经历的一些真实的、颇具代表性的问题做了一些剖析，分享我们的经验，希望对读者的学习、工作与生活能有所助益。

本书以真实的案例和数据为基础，讲述云计算和大数据知识，力求理论联系实际，深入浅出，尽量避免深奥的理论推导，语言尽可能通俗易懂。本书是一本关于云计算和大数据的科普读物。

全书共分为 5 章，分别是：揭秘云计算、揭秘大数据、云计算与大数据体系架构剖析、云计算与大数据进阶、大数据应用与云平台实战。

第 1 章揭秘云计算，着重介绍云计算简史、发展历程，与传统 IT 比较而言云计算的特质，云与业务需求的互动关系，云多重形态的存在与各自特质的介绍，剖析了不同类型云的效率博弈与比较，最后介绍了基于开源项目的云平台及服务的搭建。

第 2 章揭秘大数据，开篇介绍了大数据的前世今生，并针对当下广泛存在的对大数据的误解进行澄清，然后针对大数据所要解决的 5 大问题（大数据存储、大数据管理、大数据分析、大数据科学与大数据应用）逐一剖析，最后阐述了大数据科学的本质以及从平台与应用两个维度来分析如何构建大数据的解决方案。

第 3 章云计算与大数据体系架构剖析，首先从开源 vs. 闭源的角度阐释了整个业界的软件定义趋势、商用硬件趋势并预言了硬件回归的必然趋势，然后从 4 个层面剖析了云计算与大数据领域的技术之争——底层存储、基础架构即服务、平台即服务、应用。

第 4 章云计算与大数据进阶，给读者讲述在云与大数据的时代做什么、怎么做才是对的。内容包括靠近应用（Application Nearness）、水平可扩展（Scale-Out）、如何玩转开源（Open-Source）、怎么做服务驱动（Service-Oriented）的技术架构与运营。

第 5 章大数据应用与云平台实战，结合业界的具体实践讲解了两个平台建设的案例，一个是大数据平台的搭建，一个是混合云平台的搭建。

编者

2017 年 1 月

致 谢

笔者在过去数年间一直从事与大数据和云计算相关的前瞻性研发、走向市场 (Go-to-Market) 及内外合作交流等工作。从一开始点点滴滴累积的实验结果到工程实践的心得到四处宣讲的幻灯片再到高校教材, 资料越堆积越厚重, 久而久之就有了著书立传、答疑解惑的想法。思前想后, 笔者以为解释清楚什么是云计算+大数据最好的办法是从它们各自的来龙去脉 (寻根溯源) 入手, 于是就有了本书的初始框架——从科技发展史的角度去讲述云计算与大数据从哪里来、向何处去、本质是什么、如何认知与实现。俗话说, 万事开头难, 前面的基调定好了, 后面的细化工作反而容易了。笔者试着于 2016 年夏季学期在哈尔滨工业大学授课的时候把本书的部分内容精练汇集成教材向计算机与软件学院的本科高年级学生宣讲, 获得了不错的反响。

说到本书的书名, 尚有一个小插曲。最早在筹划时的书名笔者最中意的是《云&大数据要论》, 大抵是笔者多年从事考古、收藏, 浸淫此道日久, 不知不觉中连书名都要向此行泰斗——明代曹昭所著的《格古要论》致敬。但是, 致敬归致敬, 出版社的同仁深以为面向当代读者还是采用通俗易懂的书名为好, 于是就有了最终的书名《云计算与大数据》。

感谢在过去几年内提供素材来充实和丰富本书的朋友和同事们。我在前 EMC 研究院以及卓越研发集团跨部门的同事们: 王坤博士、曹逾博士、赵军平先生、李三平先生、董哲先生、郭晓燕先生、朱捷先生、Layne Peng 先生、Vivian Wang 女士以及 Michelle Lei 女士, 他们的很多相关科研探索及产品开发工作为本书提供了详实的数据及佐证资料。

特别需要感谢的还有我的岳父刘君胜先生和本书的责任编辑邹文波先生, 没有他们的积极策划与耐心帮助, 本书不会成为现实。

最后, 谨以此书献给我的夫人 Monica 和我们可爱的儿子 Nathan。

孙宇熙

2016 年 8 月 1 日于加州硅谷米深堂

目 录

第 1 章 揭秘云计算	1
1.1 云从哪里来	1
1.1.1 云计算科技史	1
1.1.2 业务需求推动 IT 发展	6
1.2 云的多重形态	8
1.2.1 云计算的多重服务模式	8
1.2.2 公有云 vs.私有云 vs.混合云	9
1.2.3 云的形态并非一成不变	12
1.3 关于云计算效率的讨论	18
1.3.1 公有云效率更高?	18
1.3.2 云计算优化要论	22
1.4 业界如何建云	24
1.4.1 云计算最佳实践五原则	25
1.4.2 云服务与产品的演进	28
1.4.3 开源	33
第 1 章参考文献	42
第 2 章 揭秘大数据	44
2.1 大数据从何而来?	44
2.1.1 大数据的催化剂	44
2.1.2 Data → Big Data → Data	47
2.1.3 大数据不只是 Hadoop	52
2.2 大数据的五大问题	54
2.2.1 大数据存储	55
2.2.2 大数据管理与分析	63
2.2.3 大数据科学	66
2.2.4 大数据应用	68
2.3 大数据四大阵营	71
2.3.1 OLTP 阵营	71
2.3.2 OLAP 阵营	77
2.3.3 MPP 阵营	80
2.3.4 流数据处理阵营	83
第 2 章参考文献	87

第3章 云计算与大数据体系架构剖析	89
3.1 关于开源与闭源的探讨	89
3.1.1 软件在吃所有人的午餐!	89
3.1.2 商品化硬件趋势分析	92
3.1.3 硬件回归	97
3.2 XaaS: 一切即服务	106
3.2.1 软件定义的必要性	108
3.2.2 软件定义的数据中心	113
3.2.3 软件定义的计算	117
3.2.4 软件定义的存储	121
3.2.5 软件定义的网络	126
3.2.6 资源管理、高可用与自动化	135
第3章参考文献	143
第4章 云计算与大数据进阶	144
4.1 可扩展系统构建	144
4.1.1 可扩展数据库	148
4.1.2 可扩展存储系统	153
4.2 开源模式探讨	159
4.2.1 开源业务模式	159
4.2.2 大数据开源案例	162
4.3 从SOA到MSA	165
第4章参考文献	170
第5章 大数据应用与云平台实战	172
5.1 大数据应用实践	172
5.1.1 基于开源架构的股票行情分析与预测	172
5.1.2 IMDG 应用场景	175
5.1.3 VADL (视频分析数据湖泊) 系统	178
5.2 云平台&应用实践	182
5.2.1 如何改造传统应用为云应用?	182
5.2.2 探究业界云存储平台	186
第5章参考文献	197

要搞清楚任何一个问题或一种现象，通常不外乎要从前因、后果、人、事、物的关联与分析入手，我们把以上简称为 5Ws，也就是英文中的 What-When-Where-Who-Why（什么-何时-何地-谁-为什么）的首字母提炼，也有 6Ws 的提法，多了一个 How（怎么）。对云计算的认知也不例外，我们要知道云是什么（What），云从哪里来、会到哪里去（Where-When-Who），为什么云计算在今天以至可见的未来会大行其道（Why-How）。本章会对以上相关问题做出相应的分析和说明。

1.1 云从哪里来

首先，我们需要知道云从哪里来，搞清楚谁是云计算的提出者至关重要，这个大可上升到哲学的高度，可类比千百年来科学家乃至全人类最关心的问题的核心，就是知道人从哪里来。同理，知道云从哪里来可以更好地帮助我们预判云会朝哪个方向发展，会在何处融入、改变人们的工作与生活。

1.1.1 云计算科技史

云计算的起源众说纷纭，各种版本皆有。有人说，云计算是亚马逊最早于 2006 年推出的 AWS 服务。AWS 从早期的云计算服务 EC2、存储服务 S3 到今天发展为目前业界最为广泛使用的各类计算、网络、存储、内容分发、数据库、大数据管理与应用等五花八门的服务。也有人说云计算是 Sun Microsystems 在 2006 年 3 月推出的 Sun Grid，它是一种公有云网格计算服务，一美元一小时的 CPU 使用价格，和用电一样的计费模式——Pay-Per-Use（按使用量计费）。不过，按照 MIT Technology Review 刨根问底的结果¹，Compaq（康柏）电脑公司 1996 年在内部商业计划文档（见图 1-1）中最早使用 Cloud Computing（云计算）这一字样与图标（遗憾的是，康柏在被惠普收购之后，除了继续卖了几年低端 PC 外与云计算再无瓜葛）。

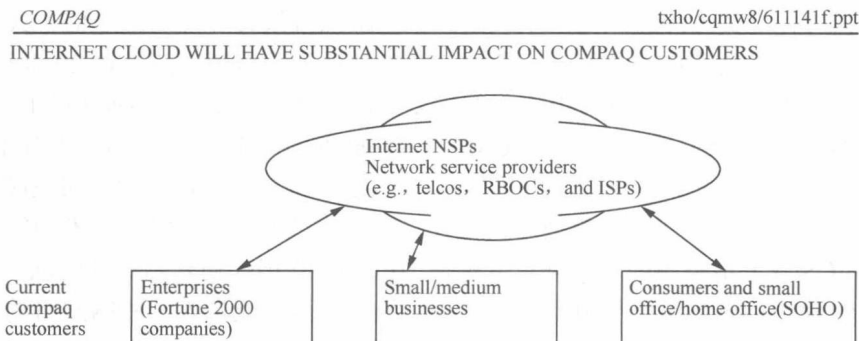


图 1-1 康柏公司 1996 年商业计划文档 ISD Strategy for Cloud Computing

1.1.1.1 云计算的三大要素

以上可以算作对云计算“冠名权”归属的一番浅究，事实上云计算的起源比以上诸多论断还要早，其发展历程贯穿了过去半个世纪全人类的 IT 发展史，如图 1-2 所示。

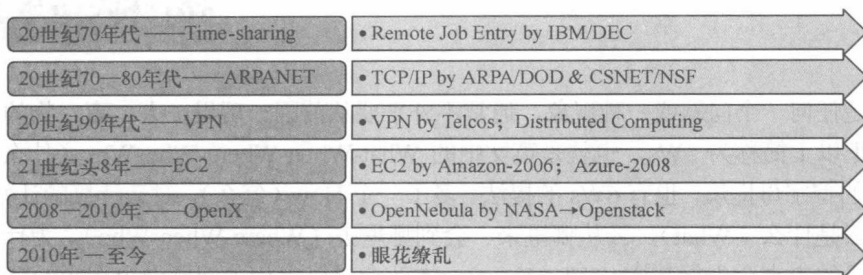


图 1-2 云的起源及发展（从 20 世纪 70 年代至今）

1. 云计算要素之一：分时计算

早在 20 世纪 70 年代初期，那还是 Mainframes（大型机）如日中天的最后一个黄金 10 年，IBM 在其大型机主机与终端之间使用了一种叫 RJE（Remote Job Entry，远程任务输入或叫远程任务处理）的机制。大型机自 20 世纪 60 年代中期就已经普遍使用了虚拟化技术，其间最典型的应用就是 Time-Sharing（分时），也就是说多个任务或者多个终端事先就被分配好占用主机处理器完成任务的时间和优先级。从整体上看，大型主机处理各任务的过程像是主机在同时服务与处理多个用户、多个终端的各种不同任务。发展到今天，从蜂窝电话网络到 PC 上面的处理器再到手机终端，分时计算在我们的生活中可谓无处不在（笔者特意没有提到虚拟化即是云计算这一容易引起误解的概念）。

2. 云计算要素之二：网络互联

从 20 世纪 60 年代末期开始到 80 年代，还有一系列大事件发生，那就是 TCP/IP 网络的诞生与蓬勃发展。这里不得不提到两张网，一个是 ARPANET [Advanced Research Projects Agency Network，（美国国防部）高级研究计划署网络]，另一个是 CSNET (Computer Science Network)。ARPANET 是美国国防部在 20 世纪 60 年代后期开始资助的用于研发 packet-switching（包交换或分组交换）技术的项目产物。ARPANET 项目产生的一个原因是，在分组交换技术之前被电信公司广泛应用的是 circuit-switching（回路交换或电路交换）技术，相比分组交换而言，回路交换使用点对点的固定通信线路对资源的利用率低下。另一个被广泛流传的 ARPANET 项目来由是当时处于美苏冷战时期，北大西洋公约组织希望有一种网络通信方式可以在经受核打击情况下依然能完成信息交换。后面这一点也被认为是美国国防部在 20 世纪 60 年代开始寻找更高效、更安全的包数据交换方式的主要原因。互联网雏形是 1969 年 12 月 5 日四所位于美国西部的高校 UCLA、Stanford、UCSB 和 UUtah 间第一次形成的四节点分组交换网络，在此前一个月最早的“互联网级”信息传递网在 UCLA 与 Stanford 大学之间完成，传送的只是一个简单的单词 login，而包交换技术仅仅在一年前才被发明（互联网技术的发展速度由此可见一斑）。CSNET 是美国国家科学基金会在 20 世纪 80 年代初开始资助的项目，可算是对 ARPANET 的有效补充（对那些由于资金或权限等限制而不能接入互联网网络的学校与机构提供帮助），此两者（ARPANET 与 CSNET）被公认为奠定了互联网的科技基石。

3. 云计算要素之三：网络安全与资源共享

20 世纪 80 年代和 90 年代，IT 行业最大的发展可以归纳为：PC 的兴起。PC 的兴起带动了整个产业链上下游的蓬勃发展，从 CPU、内存、外设到网络、存储、应用，不一而足。PC 的广泛应用极大地提升了劳动生产率，特别是在连接企业、部门、员工情景下催生了对网络资源虚拟化、数据共享以及数据传输安全的强烈需求，一系列技术应运而生，如 VPN（虚拟专网——在公共网络通道之上建立点对点的私密通信渠道）、分布式计算（PC 的发展在本质上就是让多台 PC 协同计算来实现原来只有大型机、小型机可以完成的任务）。

21 世纪的第一个 10 年，IT 行业的发展可以简化为如下两个阶段。

（1）第一阶段：2001 年—2005 年，基于 PC 的虚拟化技术高速发展。虚拟化技术早在 20 世纪 60 年代就已经在大型机上出现，不过真正广泛应用是随着 PC 的兴起，PC 单机处理能力达到了需要通过虚拟化的方式来进一步提高 PC 资源利用率。

（2）第二阶段：2006 年—2010 年，云计算服务的推出与应用。第二阶段开启于 2006 年，在同一年内 Amazon、Google、Sun Microsystems 公司先后推出了各自的云计算服务，2008 年微软公司推出了 Azure 云计算服务，同年在 NASA（美国国家航空航天局）的资助下 OpenNebula 开源云计算平台项目成立，两年后 NASA 与 Rackspace Hosting（得克萨斯州的一家 ISP）联手推出了 OpenStack 开源软件项目。在过去的数年中 OpenStack 从早期模仿 Amazon AWS 服务到逐渐形成颇具特色的 IaaS（平台即服务）平台并逐渐地向下、向上蚕食生态系统，短短几年间 OpenStack 几乎已经超越 Linux 成为全球最大的、最活跃的开源社区；而与此同时 AWS 则攻城略地，占据了公有云计算领域的大半江山。

小结以上两个阶段，可概括为：PC→PC 虚拟化→各种云计算平台及服务解决方案风起云涌。

1.1.1.2 云计算的本质

近几年，云计算的发展让人眼花缭乱，各种新兴的技术、新兴的公司风起云涌。其中值得一提的有两样东西。一个是容器计算。我们前面提过虚拟化，基于虚拟机（Virtual Machine, VM）的虚拟化可算是对 Baremetal（裸机）这种形式的有效补充，而容器可算作是对基于 VM 技术的虚拟化的有效补充。容器的意义在于重新提高了因虚拟化带来的计算效率的降低，后面的章节中我们会专门论述相关的问题。另一个值得一提的是大数据。如果说 2006 年开始的云计算浪潮多少都是偏重于底层的平台与服务，而真正寻找到的与之匹配的就是近三五年来声名鹊起的大数据应用。两者可算是一拍即合：云计算作为基础架构来承载大数据，大数据通过云计算架构与模型来提供解决方案，如图 1-3 所示。

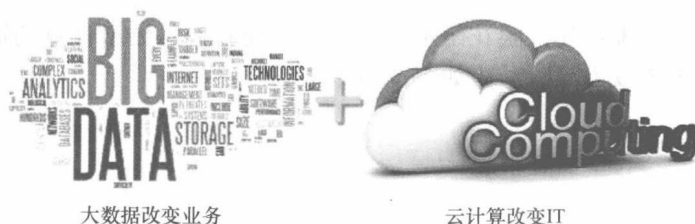


图 1-3 当云计算遇上大数据

至此，我们来总结一下到底什么是云计算。从技术角度来看，云计算是多种技术长期演变、融合的产物，诸如分布式计算、并行计算、网络存储、分布式存储、虚拟化、裸机及容器计算、负载均衡等计算机及网络技术，如图 1-4 所示。

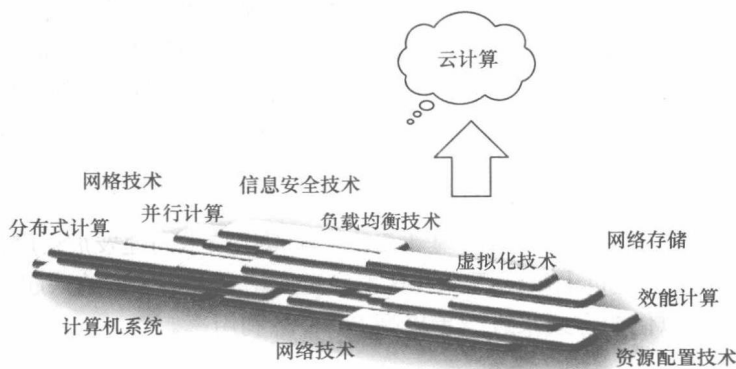


图 1-4 云计算的底层技术发展与融合

云计算的本质是多种技术的融合，它和很多其他技术颇有相通之处，例举几个。

- (1) C/S 与 B/S 技术。
- (2) P2P 技术。
- (3) 并行计算 (Parallel Computing)。

C/S (Client/Server) 泛指任何客户端到服务器端的双向通信机制和架构，而 B/S (Browser/Server)，可看作 C/S 架构的一种形式，也是目前最常见的网络节点间通信方式；P2P [Peer-to-Peer，(注意区别于 P2P 金融，后者指的是 Person-to-Person Lending，即人对人借贷，本质上是一种众筹借贷的金融模式)]，是一种分布式应用架构，它的核心理念是 Divide-n-Conquer (分而治之) 或者说是 Task-Partitioning (任务分区)。可以说 C/S 架构与任务分区理念二者合在一起就已经把云计算的技术本质描述出来了，云计算在本质上依然是一种分布式计算，一个与之相对应的技术是并行计算，它们之间的主要区别在于节点间通信方式的差异：分布式计算、云计算显然是通过信息 (Messages) 来实现节点间 (节点在这里可以理解为 CPU 或主机或计算机集群) 的通信，而并行计算通常会采用共享内存 (Shared Memory) 的方式，后者虽然效率会更高，但是可以想见其可扩展性会受到一些限制……在某种程度上，并行计算可以看作一种特殊的分布式计算，特别是在小规模紧密集群或早期分布式计算时期的实现方式中经常可以看到并行计算的影子。

1.1.1.3 云计算基本特征

这么多技术名词，难免让人眼花缭乱，那么如果要我们总结云计算的通性，或者说基本特征，有没有呢？笔者以为可总结为如下 5 条。

- (1) 共享资源池。
- (2) 快速弹性。
- (3) 可度量服务。

(4) 按需服务+自服务。

(5) 普遍的网络访问。

第一条共享资源池指的是计算、网络、存储等资源的池化和共享；第二条快速弹性指的是云计算能力应对需求、负载变化时的可伸缩性（快速弹性指的是“自动化”，原来需要几周、几个月完成的事情，现在可以在几秒钟、几分钟内完成并且几乎不需要人工干预）；第三条可度量服务也是云计算非常重要的特征，包括对各项服务和应用的监控、计费；第四条是按需服务、自服务，按需服务听起来很像某种人类文明高度发展的终极社会形态——按需分配，在云计算的背景下，按需服务是为了降低重复建设、过度分配而造成的资源浪费，而自服务则给予了用户极大的把控性同时降低了对维护成本的开销；第五条指的是可以在任何时间、任何地点通过网络来访问云计算资源。图 1-5 展示了云计算的这些基本特征。

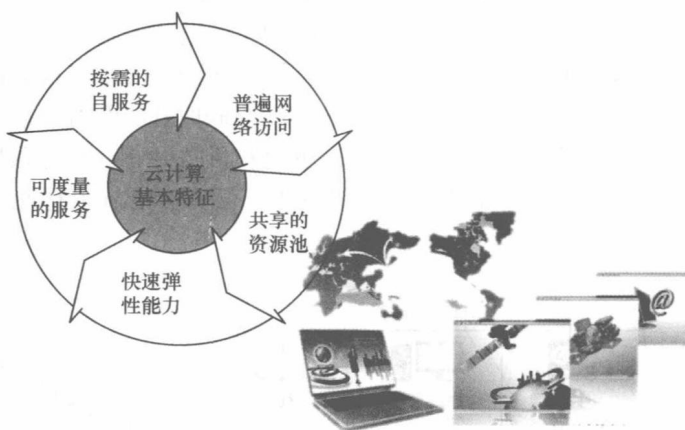


图 1-5 云计算基本特征示意图

这里，我们套用 IDC 在 2013 年提出的“三个平台²”来作为总结，如图 1-6 所示。

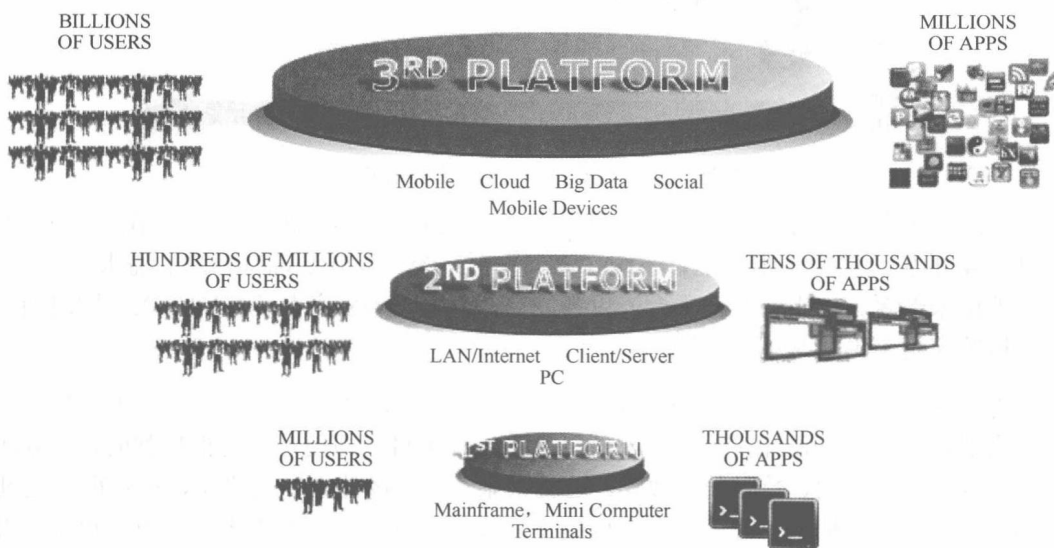


图 1-6 第一、第二、第三平台

第一平台出现最早，横跨了 20 世纪 50—70 年代，是大型机、小型机的天下，受众主要是一些大型企业，服务百万级的用户对象和数以千计的应用；第二平台是进入和连接了千家万户的 PC 的时代，用户数量与应用分别达到了亿万级和数以万计；第三平台颠覆了曾经颠覆第一平台的第二平台霸主 PC，而接入用户的规模继续以指数级增长，如果再算上 IOT 设备（物联网级产品，各种可联网传感器），则用户规模可达到数以百亿甚至千亿级，而应用也空前蓬勃发展，达到了百万量级。今天我们正处于大规模从第二平台向第三平台迁移与发展的关键阶段。在 1.1.2 节，我们会着重讨论业务需求与 IT 交付能力之间的互动问题。

1.1.2 业务需求推动 IT 发展

随着过去几十年间 IT 行业从大型主机（Mainframes）过渡到客户端服务器（PC Servers），再过渡到现如今的移动互联时代（Mobile Internet），IT 可把控的资源 and 预算的大趋势一直在下滑。在过去十几年的时间里，对虚拟化技术的采纳帮助 IT 实现了极大的效率飞跃，大幅提升了 IT 满足业务预期的能力。不过，在当下的移动互联时代，面对数以十亿计的新移动消费者以及数以百万计的新应用和服务，IT 可谓是机遇和挑战并存。业务预期呈现出了指数增长。如果不在“我们如何做 IT”方面做出根本性改变，没有人能赶上行业发展的步伐。IT 交付所面临的问题如图 1-7 所示。

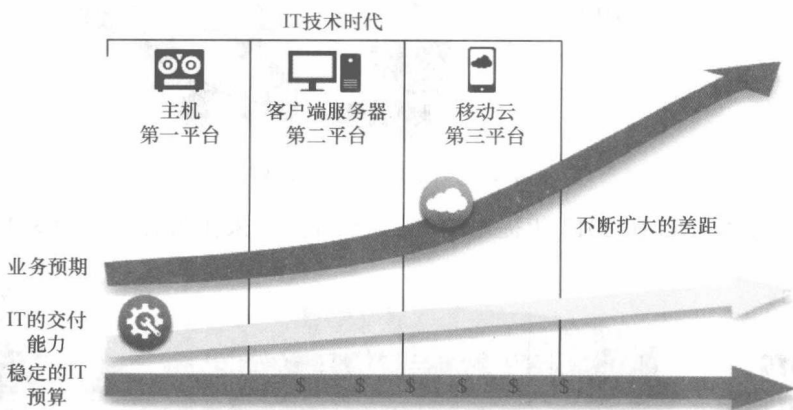


图 1-7 业务预期 vs.IT 交付能力 vs.IT 预算（横轴为时间轴）

今天绝大多数的政府和企业都已经遇到或即将遇到同样的问题，在 IT 支出中，2/3 甚至是超过 3/4 的资金被用于维护现有系统（第二平台系统）以满足现有客户的需求，仅有 1/3 的预算被用来部署新应用（第三平台）以帮助获得新客户及业务从而增加收入，同时通过大数据分析等应用来更加贴近客户等。

大多数政企 IT 部门所采用的依然是传统模型。在传统 IT 流程中，每个新的解决方案都是一个需要进行采购、设计、配置、测试以及部署的项目，即便是做得很顺利的话，新项目部署周期也要长达数周、数月甚至数年。这就很难实现高敏捷性和低投资，也就很难通过 IT 来增加收入。组织机构的目标就是增加敏捷性，减少运营支出，对未来进行更多投资，并不断降低风险，但是这两者之间存在着一个鸿沟，如图 1-8 所示。

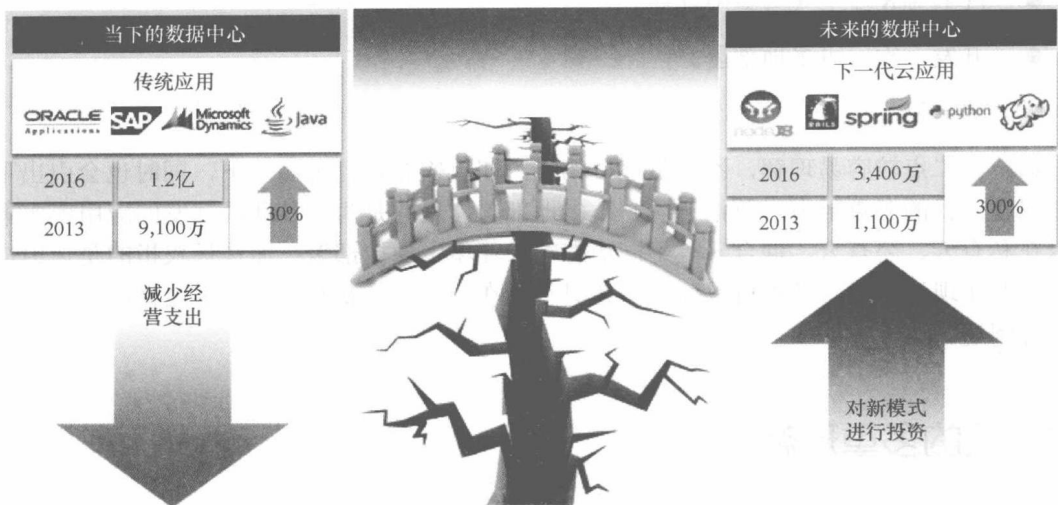


图 1-8 传统应用 vs. 下一代云应用

换一角度来看问题，今天的数据中心中依然充斥着大量的第二平台甚至第一平台的那些“传统”应用，它们虽然在增长速度上（是的，依然在增长，而不是有些人说的所有的应用都是下一代第三平台云应用，此类的说法过于绝对并且不符事实）没有新型的云应用那么惊人，但在绝对数量上依然占主体，也就是说在相当长一段可预见的时间内，政企 IT 部门的投资依然会在如何继续减少经营支出与如何增加面向新模式的投资间做出分配。

如何让 IT 的交付能力保持同步，甚至超越业务的预期是 IT 部门始终的使命（见图 1-9）。

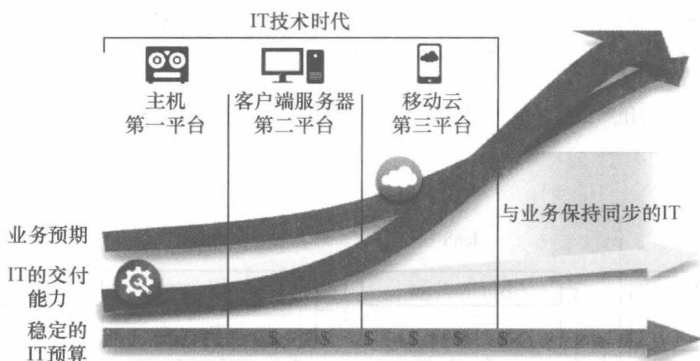


图 1-9 与业务同步的 IT 交付能力

在相当长一段时间内，业务部门对 IT 高度依赖，牺牲了敏捷性、灵活性来获得 IT 的支撑，IT 拥有极大的控制权并提供安全保障；当随着基于云的服务出现的时候，特别是三大“XaaS（即：SaaS，软件即服务；IaaS，基础架构即服务；PaaS，平台即服务）”出现后，业务部门迅速地开始拥抱它们并将之作为替代解决方案提供渠道，XaaS 带来了更高的敏捷性、灵活性，但是与此同时，IT 部门也丧失了控制，安全性也存在未知因素。走到这一步可以算是今天很多政企部门遇到的问题一个集中写照，如果继续发展下去，更高的一个阶段应该具有如下特征。

- 敏捷性。

- IT 控制&安全（可集中管理）。
- 开发一次、可多地部署。
- 选择权。

第一、二点较容易理解，不再赘述，第三、四点需要在此说明一下，同时也会引出下一节的议题：云的多重形态。所谓一次开发多地部署，指的是为一种环境开发的应用或服务，可以在私有云、公有云、混合云等不同形态的云环境上便捷部署；而选择权指的是让用户可以选择并实现让应用与数据可以跨越云边界，如在公有云、私有云间自由迁移，下一节我们会展开阐述。

1.2 云的多重形态

1.2.1 云计算的多重服务模式

云计算在快速的发展过程中逐渐形成了不同的服务模式（Service Models）。目前我们熟知的主要有三大类：SaaS（Software as a Service，软件即服务）、PaaS（平台即服务）与 IaaS（基础架构即服务），也有人喜欢把它们统称为 XaaS 或 EaaS（Everything as a Service）。类似的还有存储即服务（Storage as a Service）、容器即服务（Container as a Service）等。从根源上讲，*aaS 模式都源自 SOA（Service-Oriented Architecture），SOA 是一种架构设计模式（可类比面向对象编程语言中的设计模式），其核心就是一切以 Service 为中心，不同的应用之间通信协议都以某种服务的方式来定义和完成。今天我们经常看到的微服务架构（Micro-Service Architecture，MSA）的概念，在本质上也是由 SOA 演变而来。

为什么会形成不同的*aaS 服务模式呢？主要原因在于最终服务交付的形态，如图 1-10 所示。

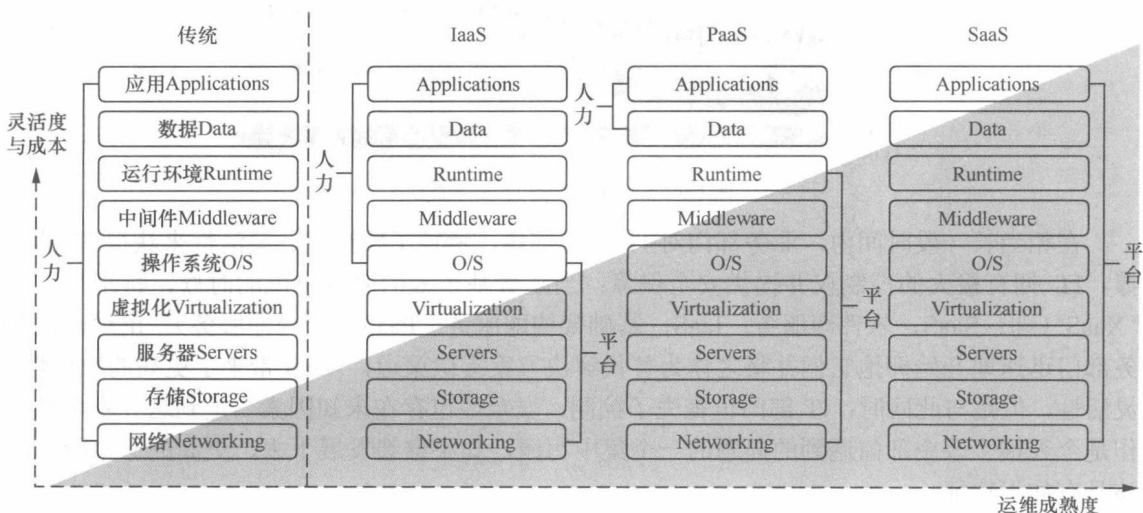


图 1-10 传统 vs. *aaS

在传统的 IT 运维与交付模式上，从最底层的各种硬件到操作系统到中间任何一层运行环境到数据与应用全要人力来维护；SaaS 是最全面的服务交付模式，从上到下所有的问题都由平台来解决，我们知道最著名的例子大概是在线 CRM 服务提供商 Salesforce.com，另一个例子是 intuit.com，从记账到报税一站式服务；IaaS 则是更多地专注于底层硬件平台与虚拟化或容器封装，而把从操作系统到上层应用的自由都留给用户来决定，最典型的就是像亚马逊或阿里云提供的云主机服务；在 SaaS 与 IaaS 中间还可能有一类服务交付模式，也是最近 1~2 年才兴起的叫作 PaaS。PaaS 可以认为是业界在看到 IaaS 交付过程和用户使用过程中遇到的各种问题后对服务交付自然延伸的必然结果，用户希望平台提供方能对操作系统、中间件、运行时甚至是应用与服务的持续升级、持续集成都提供管理，于是 PaaS 应运而生。在后面的章节中我们会专门介绍目前业界具有代表性的 PaaS 的解决方案。

在下面的小节中，我们先来了解云的不同部署形态。

1.2.2 公有云 vs.私有云 vs.混合云

从云架构部署、服务、应用以及访问的方式来区分，我们一般把云分为四大类：

- 私有云；
- 公有云；
- 混合云；
- 社区云。

前两朵云耳熟能详，混合云顾名思义，既包含私有云，又包含公有云。最后的社区云相对少见，列在这里是为了完整起见，它可能是一种特殊形态的私有云、公有云或混合云，通常由一些协作的组织在一定时间内基于同样的业务诉求与安全需求而构建形成的一种云架构。下面我们主要谈谈对前三朵云的认知。

对私有云、公有云、混合云的定义与界定的关键是云的服务对象。

- 如果被服务的对象是一个机构，那么我们称之为私有云。
- 如果服务开放给大众，并通过互联网可以访问，则称之为公有云。
- 混合云通常是兼有私有云和公有云部分，两部分可相对独立运作，但是也会协同工作，例如一些任务可能会横跨公有云、私有云边界。

大家还会经常听到专有云的提法，在本质上这依然是私有云的一种，它经常被托管的私有云的方式存在，目前各地方政府和企业经常会把自己的一部分可以放入云中或向云端迁移的业务托管给第三方云服务提供商来运维。

从“物种起源”的角度上说公有云与私有云都是由早年的 DC/IDC 数据中心或互联网数据中心发展而来。除了它们各自所可能侧重的服务不同外，在技术本质上没有高低优劣之分（见图 1-11）。