

復旦大學

云计算领域调研报告

王傲

15300240004

管理信息系统

SOFT130068.01

指导教师：朱东来

## 目录:

0. 目录 .....	2
1. 云计算基本情况 .....	3
1.1 云计算的概念 .....	3
1.2 云计算的发展历程 .....	4
2. 云计算的分类 .....	5
2.1 IaaS 模式 .....	5
2.1.1 IaaS 模式发展现状 .....	5
2.1.2 IaaS 模式产品及应用 .....	6
2.2 PaaS 模式 .....	7
2.2.1 PaaS 模式发展现状 .....	7
2.2.2 PaaS 模式产品及应用 .....	7
2.3 SaaS 模式 .....	8
2.3.1 SaaS 模式发展现状 .....	8
2.3.2 SaaS 模式产品及应用 .....	8
3. 云计算关键技术 .....	9
3.1 虚拟化技术 .....	10
3.2 分布式数据存储技术 .....	10
3.3 大规模资源管理技术 .....	12
3.4 并行编程技术 .....	13
3.5 其他关键技术 .....	14
4. 云计算的商业模式及发展 .....	15
4.1 云计算的商业模式 .....	15
4.2 国外云计算商业发展状况 .....	16
4.3 国内云计算商业发展状况 .....	19
5. 未来展望 .....	21
6. 结论 .....	21
7. 参考资料 .....	21

## 内容摘要:

本次调研对云计算领域进行了比较详细的调查。调研对云计算的现状进行了基本概述，对云计算进行了分类并介绍了相关主要产品，展示了云计算的主要技术，然后针对云计算的商业模式对国内外龙头企业和云计算产业链进行了调查，最终给出了结论。

**关键字：**云计算 IaaS PaaS SaaS GFS BigTable MapReduce 亚马逊 微软 Google 阿里

## 1. 云计算基本情况

### 1.1 云计算的概念

云计算（Cloud Computing），是一种新兴的共享基础构架的方法，可以将巨大的系统池连接在一起为用户提供各种 IT 服务。云计算被很多人视为“革命性的计算模型”，因为它使得超级计算能力通过互联网自由流通成为了可能。由于云计算是一个概念，而不是指某项具体的技术或标准，所以不同的人从不同的角度出发就会有不同的理解。现阶段广为接受的是美国国家标准与技术研究院（NIST）的定义：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（包括网络、服务器、存储、应用软件、服务等），这些资源能够被快速提供，只需投入很少的管理工作或服务供应商进行很少的交互。

结合一些应用场景解释，云，可以理解为一系列的计算资源，通过网络提供服务。比如苹果公司的 iCloud, 各家公司的网盘等等，可以把资料照片备份在云端，也就是远端服务器，即使更换电脑或者更换手机也能从云端获取曾经备份的资料和照片，这是最简单的云计算应用场景。

更具普遍意义的讲，云计算通常是指各种计算资源的组合，比如 Google，全球有超百万台服务器来处理用户的搜索请求，这些服务器可以理解为一个云，用户不需要关心云端的细节，每一次搜索可能会调用无数台云端服务器进行搜索，最后汇总结果展现给客户，这也是一种云服务。再底层一些的面向开发者的云计算服务，比如新浪、Facebook 提供的开放 API，针对厂商 API 开发的程序可以部署在云端，这些 API 提供了对底层硬件资源的访问途径，开发者可以利用这些服务实现自己的业务逻辑。

又比如阿里云、腾讯云，或者亚马逊 EC2，采用虚拟化技术，用户只需要选定自己需要的硬件资源，比如硬盘、内存、带宽、操作系统，就立即动态生成一台用户看上去仿佛

物理唯一的机器，免去了自己购买服务器或者带宽的麻烦，使得业务能迅速建立，并且动态扩展。

## 1.2 云计算的发展历程

云计算是多种技术混合演进的结果，其成熟度较高，又有大公司推动，发展极为迅速。Google、亚马逊、IBM、微软和 Yahoo 等大公司云计算的先行者。云计算领域的众多成功公司还包括 VMware、Salesforce、Facebook、MySpace 等。

亚马逊在云计算领域起步非常早，并且在全球占有非常大的市场份额。亚马逊最早研发了弹性计算云 EC2 (Elastic Computing Cloud) 和简单存储服务 S3 (Simple Storage Service) 为企业计算和存储服务。在诞生不到两年的时间内，亚马逊的注册用户就多达 44 万人，其中包括为数众多的企业级用户。亚马逊的 AWS 自 2006 年推出 S3 和 EC2 后服务至今，历经 10 年，2015 年营收 78.8 亿美元，其市场份额大约在 30% 左右。有研究机构认为 AWS 的计算能力超过其他 14 家云服务商（包括微软、Google、IBM 在内）之和。

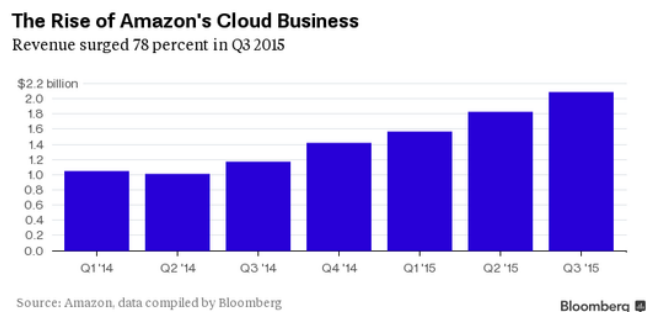


图 1 亚马逊 AWS 15 年第三季度之前的营收额对比

Google 是最大的云计算技术的使用者。Google 搜索引擎就建立在分布于 30 多个站点、超过 200 万台的服务器的基础上，而且这些设施的数量正在迅猛增长。Google 的一系列成功应用平台，包括 Google Earth、Google Map、Gmail、Google Scholar 等也同样使用了这些基础设施。使用 Google Docs 之类的应用，用户数据会保存在互联网上的某个位置，可以通过任何一个与互联网相连的终端十分便利地访问和共享这些数据。这是云的应用之一。

IBM 在 2007 年 11 月推出了“改变游戏规则”的“蓝云”计算平台，为客户带来即买即用的云计算平台。它包括一系列自我管理和自我修复的虚拟化云计算软件，使来自全球

的应用可以访问分布式的大型服务器池，使得数据中心在类似于互联网的环境下运行计算。IBM 正在与 17 个欧洲组织合作开展以“无障碍的资源和服务虚拟化”为口号的名为 RESERVOIR 的云计算项目，并获得了欧盟提供的 1.7 亿欧元作为部分资金。

微软紧跟云计算步伐，于 2008 年 10 月推出了 Windows Azure 操作系统。Azure（译为“蓝天”）是继 Windows 取代 DOS 之后，微软的又一次颠覆性转型——通过在互联网架构上打造新云计算平台，让 Windows 真正由 PC 延伸到“蓝天”上。Azure 的底层是微软全球基础服务系统，由遍布全球的第四代数据中心构成。目前，微软已经配置了 220 个集装箱式数据中心，包括 44 万台服务器。

在我国，云计算发展也非常迅猛。2008 年，IBM 先后在无锡和北京建立了两个云计算中心；世纪互联推出了 CloudEx 产品线，提供互联网主机服务、在线存储虚拟化服务等；中国移动研究院已经建立起拥有 1024 个服务器的云计算试验中心；解放军理工大学研制了云存储系统 MassCloud，并以它支撑基于 3G 的大规模视频监控应用和数字地球系统。

目前，在国内，云计算方面占据市场份额最大的是阿里云。国内云计算厂家并不算多，却在急速增长，但基本上是一超多强的局面：阿里云占据了超过 50% 的市场，腾讯云、金山云、华为云等同属第二梯队，距离阿里云有较大的差距。阿里云目前在弹性计算领域的积累已经可比肩国际巨头 AWS、Azure，并且在国外市场也得到了认可。

## 2. 云计算的分类

云计算可以分为 IaaS, PaaS, SaaS 三种模式。

### 2.1 IaaS 模式

#### 2.1.1 IaaS 模式发展现状

IaaS (Infrastructure as a Service, 基础设施即服务), 可以利用完善的计算机基础设施, 通过互联网向消费者提供服务。IaaS 通过互联网提供了数据中心、基础架构硬件和软件资源。IaaS 还可以提供服务器、操作系统、磁盘存储、数据库和信息资源。最高端 IaaS 的代表产品是亚马逊的 AWS (Elastic Compute Cloud, EC2), 不过 IBM、VMware 和惠普以及其他一些传统 IT 厂商也提供这类的服务。IaaS 通常会按照“弹性云”的模式引入其他的使用和计价模式, 也就是在任何一个特定的时间, 都只使用用户需要的服务, 并且只为之付费。

IaaS 将硬件设备等基础资源封装成服务供用户使用，在 IaaS 环境中，用户相当于在使用裸机和磁盘，既可以让它运行 Windows，也可以让它运行 Linux，因而几乎可以做任何想做的事情，但用户必须考虑如何才能让多台机器协同工作起来。AWS 提供了在节点之间互通消息的接口简单队列服务 SQS (Simple Queue Service)。IaaS 最大优势在于它允许用户动态申请或释放节点，按使用量计费。运行 IaaS 的服务器规模达到几十万台之多，用户因而可以认为能够申请的资源几乎是无限的。而 IaaS 是由公众共享的，因而具有更高的资源使用效率。

IaaS 又分为两种：公共的和私有的。亚马逊 EC2 在基础设施云中使用公共服务器池。更加私有化的服务会使用企业内部数据中心的一组公共或私有服务器池。如果在企业数据中心环境中开发软件，那么这两种类型都有可能使用，而且使 EC 临时扩展资源的成本也很低。将两者结合起来使用可以更快的开发应用程序和服务，缩短开发和测试周期。

IaaS 服务商领域鱼龙混杂，既有多年从事 IDC 服务的专业服务商，也有一些规模较小的只能提供一些简单的虚拟化服务的厂商。IaaS 服务商大致可以分成五大类：第一类，既能提供 IaaS 解决方案，又能提供 IaaS 服务的软硬件厂商，比如 IBM、惠普、VMware 等；第二类，拥有良好基础设施的互联网厂商，比如亚马逊、Google 等；第三类，在资金、客户、IDC 方面颇具实力的电信运营商，比如中国移动、中国电信、AT&T 等；第四类，一直从事数据中心外包服务的厂商，比如万国数据、中金数据、云快线等；第五类，政府投资建设的开发区或软件园区的云数据中心。

总体来看，中国 IaaS 服务市场还处于发展初期，整体规模不大，竞争也不是非常激烈。IaaS 服务市场在快速发展的同时，还应该不断提升 IaaS 技术的成熟度，在商业模式上实现创新。

### 2.1.2 IaaS 模式产品及应用

亚马逊在 IaaS 服务领域的成功被奉为经典，而它的成功源自于它的战略眼光、对市场的深入了解以及在技术上的积累和商业模式上的大胆创新。亚马逊拥有现成的、庞大而完善的基础设施，在提供 IaaS 服务时不必再重新对基础设施进行投入。这些都是亚马逊投身 IaaS 服务市场的天然优势。亚马逊所提供的 EBS 服务，属于云硬盘，而提供的 S3 服务，属于云存储。云存储是一种 Key/Value 数据存储服务，用户通过 API

进行数据存取操作，也可以通过 S3FS 等手段将云存储作为虚拟硬盘同时挂载在多台云主机上。终端用户访问云存储的数据，不需要经过云主机。

除了亚马逊,这个领域还充满了来自其他厂商的产品，比如业界的第一个解决方案，IBM 的 Blue Cloud（蓝云）；Cisco 的 UCS 和 Joyent 的基于 IaaS 的服务等。

## 2.2 PaaS 模式

### 2.2.1 PaaS 模式发展现状

PaaS (Platform as a Service, 平台即服务) 对资源的抽象层次更进一层，它提供用户应用程序的运行环境，典型的如 Google App Engine。微软的云计算操作系统 Microsoft Windows Azure 也可大致归入这一类。PaaS 自身负责资源的动态扩展和容错管理，用户应用程序不必过多考虑节点间的配合问题。但与此同时，用户的自主权降低，必须使用特定的编程环境并遵照特定的编程模型。这有点像在高性能集群计算机里进行 MPI 编程，只适用于解决某些特定的计算问题。例如，Google App Engine 只允许使用 Python 和 Java 语言、基于 Django 的 Web 应用框架、调用 Google App Engine SDK 来开发在线应用服务。

PaaS 实际上是指将软件研发的平台作为一种服务，以 SaaS 的模式提交给用户。因此，PaaS 也是 SaaS 模式的一种应用。但是，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，因为它能够为企业提供定制化研发的中间件平台，同时涵盖数据库和应用服务器等。PaaS 可以提高在 Web 平台上利用的资源数量。用户或者厂商基于 PaaS 平台可以快速开发自己所需要的应用和产品。同时，PaaS 平台开发的应用能更好地搭建基于 SOA 架构的企业应用。

PaaS 运营商所需提供的服务，不仅仅是单纯的基础平台，而且包括针对该平台的技术支持服务，甚至针对该平台而进行的应用系统开发、优化等服务。

### 2.2.2 PaaS 模式产品及应用

比较知名的 PaaS 平台包括 Google App Engine、Salesforce、Windows Azure、Amazon AWS 等等。

GAE (Google App Engine) 是一个开发、托管网络应用程序的平台，使用 Google 管理的数据中心。GAE 应用程序易于构建和维护，并可根据用户的访问量和数据存储需要的增长轻松扩展。通过 GAE，用户可以在支持 Google 应用程序的相同系统上构建和承载网络

应用程序。GAE 可提供快速开发和部署功能，管理简单，无需担心硬件、补丁或备份，并可轻松实现可扩展性。

Elastic Beanstalk 为在 Amazon Web Services 云中部署和管理应用提供了一种方法。该平台建立如面向 PHP 的 Apache HTTP Server 和面向 Java 的 Apache Tomcat 这样的软件栈。开发人员保留对 AWS 资源的控制权，并可以部署新的应用程序版本、运行环境或回滚到以前的版本。CloudWatch 提供监测指标，如 CPU 利用率、请求计数、平均延迟等。通过 Elastic Beanstalk 部署应用程序到 AWS，开发人员可以使用 AWS 管理控制台、Git 和一个类似于 Eclipse 的 IDE。

## 2.3 SaaS 模式

### 2.3.1 SaaS 模式发展现状

SaaS (Software as a Service, 软件即服务) 是最为成熟、最出名，也是得到最广泛应用的一种云计算服务。可以将它理解为一种软件分布模式，在这种模式下，应用软件安装在厂商或者服务供应商那里，用户可以通过某个网络来使用这些软件(通常使用的网络是互联网)。这种模式通常也被称为“按需应变”(on demand)软件。这是最成熟的云计算模式，因为这种模式具有高度的灵活性、已经证明可靠的支持服务、强大的可扩展性，因此能够降低客户的维护成本和投入，而且由于这种模式的多宗旨式的基础架构，运营成本也得以降低。SaaS 供应商通常是按照客户所租用的软件模块来进行收费的，因此用户可以根据需求按需订购软件应用服务，而且 SaaS 的供应商会负责系统的部署、升级和维护。而传统管理软件通常是买家需要一次支付一笔可观的费用才能正式启动。

ERP 这样的企业应用软件，软件的部署和实施比软件本身的功能、性能更为重要，万一部署失败，所有的投入几乎全部白费，这样的风险是每个企业用户都希望避免的。通常的 ERP、CRM 项目的部署周期至少需要一两年甚至更久的时间，而 SaaS 模式的软件项目部署最多也不会超过 90 天，而且用户无需在软件 Licence 和硬件方面进行投资。传统软件在使用方式上受空间和地点的限制，必须在固定的设备上使用，而 SaaS 模式的软件项目可以在任何时间在任何可接入互联网的地方使用。相对于传统软件而言，SaaS 模式在软件的升级、服务、数据安全传输等各个方面都有很大的优势。

### 2.3.2 SaaS 模式产品及应用



阿里巴巴 2007 年刚刚开始就宣布进军软件行业，要通过 SaaS 的模式提供软件服务；Google 在 2006 年发布了自己的在线 Office，正式发起对微软 Office 的挑战；此外，还有 Google Apps、Salesforce 的 CRM 等等。越来越多的软件企业开始通过互联网提供软件服务，同时越来越多的互联网用户也加入到了 SaaS 服务中来，加速了软件与互联网的融合。互联网给企业带来的不仅仅是发布和收集信息的窗口，更为企业的数据管理和信息应用提供了丰富的信息资源。SaaS 把这些信息管理起来，成就了互联网和软件的融合。

伴随 SaaS 发展起来的在线 OS、在线播放器、在线 IM、在线 OA 等软件应用将越来越多，低廉的使用价格也可以更好的打击盗版。越来越多低使用成本的 SaaS 应用逐渐覆盖了传统 Licence 模式的软件，也让盗版市场无利可图。另外，SaaS 改变了传统软件的购买 Licence 使用模式，转变为租用服务，让企业的使用成本降到最低，大大降低了中小企业信息化的实施门槛。

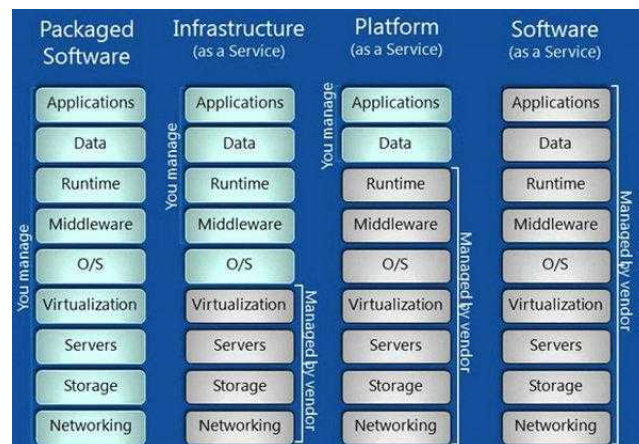


图 2 IaaS, PaaS, SaaS 的对比

需要指出的是，随着云计算的深化发展，不同云计算解决方案之间会相互渗透融合，同一种产品往往横跨两种以上类型。例如，Amazon Web Services 是以 PaaS 起家的，新提供的弹性 MapReduce 服务模仿了 Google 的 MapReduce，简单数据库服务 SimpleDB 模仿了 Google 的 BigTable，这二者仍属于 PaaS 的范畴，但它新提供的电子商务服务 FPE 和 DevPay 以及网站访问统计服务 Alexa Web 服务，则属于 SaaS 的范畴。

### 3. 云计算关键技术

### 3.1 虚拟化技术

虚拟化是云计算中主要支撑技术之一。云计算的虚拟化技术不同于传统的单一虚拟化，它是涵盖整个 IT 架构的，包括资源、网络、应用和桌面在内的全系统虚拟化。它的优势在于能够把所有硬件设备、软件应用和数据隔离开来，打破硬件配置、软件部署和数据分布的界限，实现 IT 架构的动态化，实现资源集中管理，使应用能够动态地使用虚拟资源和物理资源，提高系统适应需求和环境的能力。虚拟化允许 IT 部门添加、减少移动硬件和软件到它们想要的地方。虚拟化为组织带来灵活性，从而改善 IT 运维和减少成本支出。一旦接受云计算作为总方针来运行业务，通过简化管理流程和提高效率来降低总成本可以为虚拟化平台带来巨大的价值。目前云计算，大多是依赖虚拟化，通过把多台服务器实体虚拟化后，构成一个资源池，实现共同计算，共享资源。

而对于信息系统仿真，云计算虚拟化技术的应用意义不仅仅在于提高资源利用率并降低成本，更大的意义是提供强大的计算能力。众所周知，信息系统仿真系统是一种具有超大计算量的复杂系统，计算能力对于系统运行效率、精度和可靠性影响很大，而虚拟化技术可以将大量分散的、没有得到充分利用的计算能力，整合到计算高负荷的计算机或服务器上，实现全网资源统一调度使用，从而在存储、传输、运算等多个方面达到高效。

云计算和虚拟化是密切相关的，但是虚拟化对于云计算来说并不是唯一的支柱功能。云计算为基础设施带来的服务如：管理一个私有云（在数据中心）、公共云（比如 Salesforce）和管理托管云（托管在别处的虚拟服务器）以及许多其他的增值服务，这些都是虚拟化和云计算的不同。云计算将各种 IT 资源以服务的方式通过互联网交付给用户，然而虚拟化本身不能给用户提供服务层，而没有自服务层，就不能提供计算服务。

虽然虚拟化和云计算并非是捆绑技术，但二者可以通过优势互补为用户提供更优质的服务。云计算方案使用虚拟化技术使整个 IT 基础设施的资源部署更灵活。反过来，虚拟化方案也可以引入云计算的理念，为用户提供按需使用的资源和服务。在一些特定业务中，云计算和虚拟化是分不开的，只有同时应用这两项技术，服务才能顺利开展。

### 3.2 分布式数据存储技术

前面提到，云计算的优势之一就是能够快速、高效地处理海量数据。在数据爆炸的今天，这一点至关重要。为了保证数据的高可靠性，云计算通常会采用分布式存储技术，将

数据存储在不同的物理设备中。这种模式不仅摆脱了硬件设备的限制，同时扩展性更好，能够快速响应用户需求的变化。

分布式存储与传统的网络存储并不完全一样，传统的网络存储系统采用集中的存储服务器存放所有数据，存储服务器成为系统性能的瓶颈，不能满足大规模存储应用的需要。分布式网络存储系统采用可扩展的系统结构，利用多台存储服务器分担存储负荷，利用位置服务器定位存储信息。它不但提高了系统的可靠性、可用性和存取效率，还易于扩展。在当前的云计算领域，Google 的 GFS 和 Hadoop 开发的开源系统 HDFS 是比较流行的两种云计算分布式存储系统。

以 Google 的“三驾马车”之一 GFS（Google File System）为例。GFS 分为 Client（客户端）、Master（主服务器）和 Chunk Server（数据块服务器）。Client 是 GFS 提供给应用程序的访问接口；Master 是 GFS 的管理节点，保存系统的元数据，负责整个文件系统的管理，是 GFS 文件系统上的“大脑”；Chunk Server 负责具体的存储工作。客户端访问 GFS 时，首先访问 Master 节点，获取与之进行交互的 Chunk Server 信息，然后直接访问这些 Chunk Server，完成数据存取工作。GFS 实现了控制流和数据流的分离：Client 与 Master 之间只有控制流，而无数据流，极大地降低了 Master 的负载。Client 与 Chunk Server 之间直接传输数据流，由于文件分成多个 Chunk 分布式存储，Client 可以同时访问多个 Chunk Server，从而使得整个系统的 I/O 高度并行，系统整体性能得到提高。

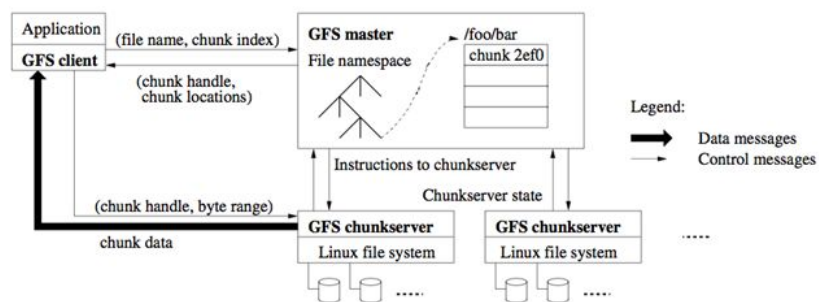


Figure 1: GFS Architecture

图 3 GFS 架构

云计算系统所处理的资源往往非常庞大，少则几百台服务器，多则上万台，同时可能跨跃多个地域，且云平台中运行的应用也是数以千计。如何有效地管理这批资源，保证它

们正常提供服务，需要强大的技术支撑。由此可见以 Google 的 GFS 和 Hadoop 的 HDFS 为代表的分布式存储系统的重要性。

### 3.3 大规模资源管理技术

信息系统仿真系统在大多数情况下会处在多节点并发执行环境中，要保证系统状态的正确性，必须保证分布数据的一致性。为了解决分布一致性问题，计算机界的很多公司和研究人员提出了各种各样的协议，这些协议是一些需要遵循的规则，也就是说，在云计算出现之前，解决分布一致性问题是靠众多协议来实现的。但对于大规模，甚至超大规模的分布式系统来说，无法保证各个分系统、子系统都使用同样的协议，也就无法保证分布一致性问题得到解决。

云计算中的大规模资源管理技术圆满解决了这一问题，而这是以上面的分布式数据存储技术为基础的。云计算系统对大数据集进行处理、分析并向用户提供高效的服务。因此，数据管理技术必须能够高效的管理大数据集。其次，如何在规模巨大的数据中找到特定的数据也是云计算数据管理技术所必须解决的问题。云计算的特点是对海量的数据存储、读取后进行大量的分析，数据的读操作频率远大于数据的更新频率，所以云中的数据管理是一种读优化的数据管理。因此，云系统的数据管理往往采用数据库领域中列存储的数据管理模式，将表按列划分后存储。

为了满足云计算的大规模数据集管理和高效的数据定位需求，不同公司采用了不同的数据管理解决方案。Google 的 BT (BigTable) 数据管理技术和 Hadoop 团队开发的开源数据管理模块 HBase 是业界比较典型的大规模数据管理技术。以 Google 的另一辆“三驾马车”之一 BigTable 为例。BigTable 数据管理方式的设计者 Google 给出了如下定义：BigTable 是一种为了管理结构化数据而设计的分布式存储系统，这些数据可以扩展到非常大的规模，例如在数千台商用服务器上的达到 PB (Petabytes) 规模的数据。BigTable 对数据读操作进行优化，采用列存储的方式，提高数据读取效率。BigTable 管理数据的存储结构为：< row : string, column : string, time : int64 > --> string:

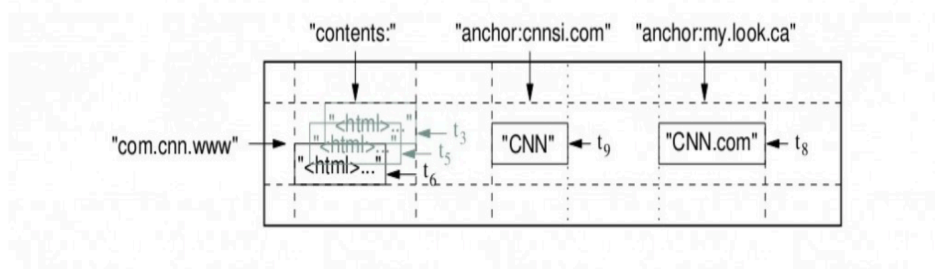


Figure 1: A slice of an example table that stores Web pages. The row name is a reversed URL. The contents column family contains the page contents, and the anchor column family contains the text of any anchors that reference the page. CNN's home page is referenced by both the Sports Illustrated and the MY-look home pages, so the row contains columns named anchor:cnnsi.com and anchor:my.look.ca. Each anchor cell has one version; the contents column has three versions, at timestamps  $t_3$ ,  $t_5$ , and  $t_6$ .

图 4 BigTable 的存储数据模型

BigTable 的基本元素是行，列，记录板和时间戳。其中，记录板是一段行的集合体。BigTable 中的数据项按照行关键字的字典序排列，每行动态地划分到记录板中。每个节点管理大约 100 个记录板。时间戳是一个 64 位的整数，表示数据的不同版本。BigTable 在执行时需要三个主要的组件：链接到每个客户端的库，一个主服务器和多个记录板服务器。主服务器用于分配记录板到记录板服务器以及负载平衡，垃圾回收等。记录板服务器用于直接管理一组记录板，处理读写请求等。为保证数据结构的高可扩展性，BigTable 采用一个类似 B+树的三级层次化的方式来存储位置信息。其中第一级的 Chubby file 中包含 Root Tablet 的位置，Root Tablet 包含所有 METADATA tablets 的位置信息，每个 METADATA tablets 包含许多 User Table 的位置信息。

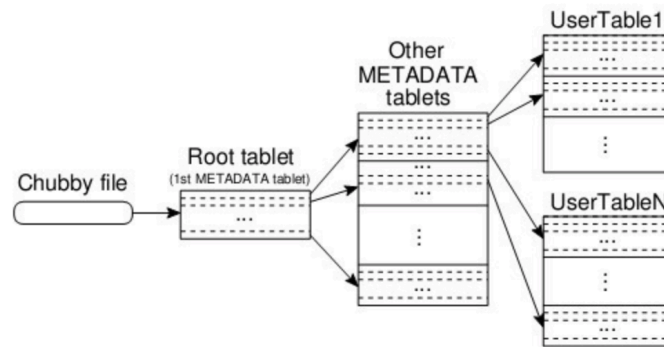


图 5 BigTable 的三级层次存储结构

而像 BigTable 这样的大规模资源管理技术，保证了云计算中面对海量数据时的准确性和效率，提高了云计算处理数据的准确性和可靠性的同时也降低了成本。

### 3.4 并行编程技术

云计算采用并行编程模式。在并行编程模式下，并发处理、容错、数据分布、负载均衡等细节都被抽象到一个函数库中，通过统一接口，用户大尺度的计算任务被自动并发和分布执行，即将一个任务自动分成多个子任务，并行地处理海量数据。

分布式并行编程模式创立的初衷是更高效地利用软、硬件资源，让用户更快速、更简单地使用应用或服务。在分布式并行编程模式中，后台复杂的任务处理和资源调度对于用户来说是透明的，这样用户体验能够大大提升。MapReduce 是当前云计算主流并行编程模式之一。MapReduce 模式将任务自动分成多个子任务，通过 Map 和 Reduce 两步实现任务在大规模计算节点中的调度与分配。

作为 Google 的“三驾马车”中的最后一辆, MapReduce 是 Google 开发的编程模型，主要用于大规模数据集（大于 1TB）的并行运算。MapReduce 模式的思想是将要执行的问题分解成 Map（映射）和 Reduce（化简）的方式，先通过 Map 程序将数据切割成不相关的区块，分配（调度）给大量计算机处理，达到分布式运算的效果，再通过 Reduce 程序将结果汇整输出。

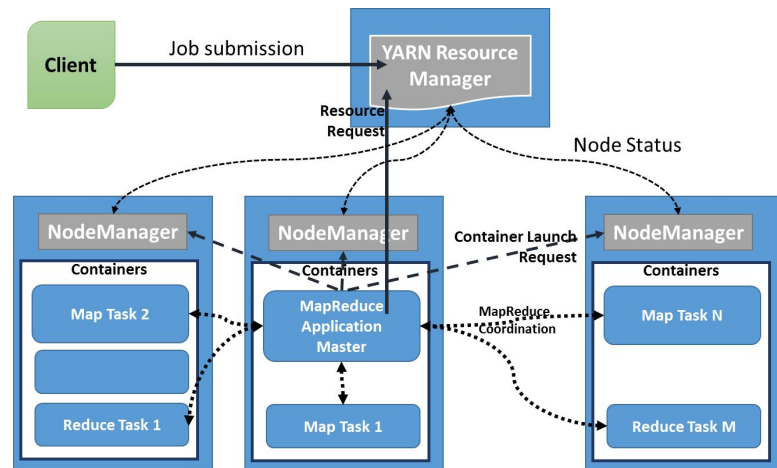


图 6 MapReduce 架构

从实际来看，MapReduce 这样的并行编程模式架构十分成功，非常适合在云计算中更高效的利用软硬件资源处理大规模海量数据，成为云计算的又一技术基石。

### 3.5 其他关键技术

云计算中使用的其他关键技术还有很多，比如云计算中的信息安全技术、大规模云计算平台管理、绿色节能技术等。这些关键技术一起，为云计算的普及和发展奠定了基础，而云计算中面临的问题，也需要发展新的技术来解决。



## 4. 云计算的商业模式及发展

### 4.1 云计算的商业模式

从云计算最初定义可以看出，云计算与传统 IT 产业的区的区别在于其商业卖点的区别：传统 IT 产业主要是以卖软硬件产品为主，而卖的方式主要以物理实际量为单位，通过传统渠道进行销售；但由于用户的需求的日益多样与复杂，也使得传统 IT 产业的产品与服务逐渐密不可分，互为载体，相互支撑；产品与服务之间的边界日趋融合与模糊，二者之间的一体化程度明显提高。以往作为偏重产品层面的云计算则应该在服务层面实现更多的发展与突破。除此之外，由于成本、功能等方面的考虑，传统 IT 产业也呈现软硬件产品一体化的趋势，这一点在云计算中将会发挥到极致。因此，以服务为导向的云计算作为新业务脱颖而出。并且，随着苹果、Google 进军电讯行业，电讯厂商也开始进入 IT 行业，ICT 之间的融合已经成为新的发展趋势。从应用角度来看，随着移动互联网、物联网、车联网、社交网等的快速发展，产生了难以计量的大数据，而云计算也成为大数据计算和处理、存储的重要手段。因此，数据和应用应该成为未来云计算厂商的关键业务构成。现在，在我国发展最快的是 IaaS 和 SaaS，在 PaaS 方面还是比较弱。但是，就 PaaS 来说，对于云计算厂商而言，是应该去努力发展的重要方向。从当前领域来看，PaaS 可以作为云计算厂商自己使用的工具推动其它两大层面的发展。

云计算和传统 IT 行业并不是对立的，而是相辅相成的：IT 行业技术的发展推动了云计算的发展，而云计算则可以为 IT 行业提供成本更低的基础服务。

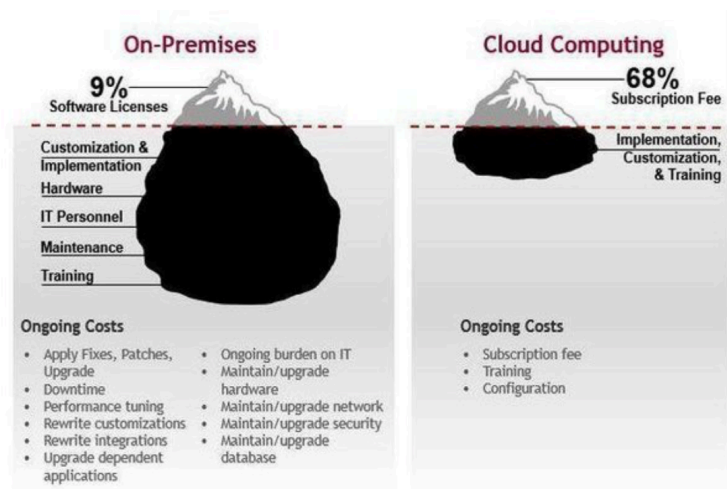


图 7 云计算可以显著降低 IT 行业的隐性成本

从上面这张图可以看到，IT 的成本其实分成很多部分，比如硬件成本、IT 维护人员成本、IT 的培训成本、项目实施开发成本等，而这些成本是隐性的。很多人可能只看到了硬件成本，或者是购买软件 Licence 的成本，容易忽略人员成本、培训成本、开发成本等很多其他的成本。云计算将大部分的成本暴露在水面之上，并且把大量前期投入的固定成本转化为运营成本。另外，随着社会的发展，人员成本会越来越高，而通过集约化的方式，云计算服务商所提供的服务，其实已经覆盖了大部分的基础运维（安装系统、配置网络、安全、监控等），让客户只需要专注于应用层的运维。可以预见，云计算在这里节约的费用，将会越来越多。另外，对于创业公司，最宝贵的其实是时间，时间成本同样不容忽视。比如在直播业务竞争白热化的今天，缩减产品的开发上线周期，将会建立非常大的竞争优势，而这里的价值，是无法用金钱来衡量的。云计算，就可以做到降低这部分成本，大大提高产品部署效率的同时节约了很大部分的成本，提高了竞争力，而这是云计算商业模式成功的重要原因之一。

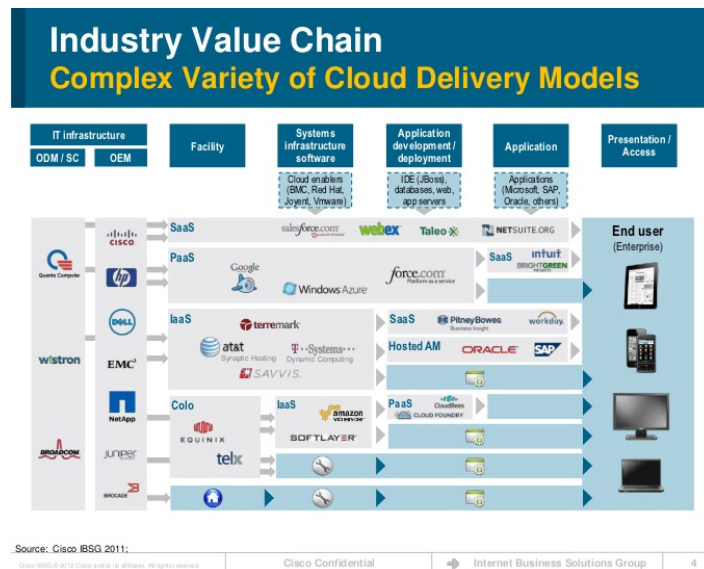


图 8 云计算产业链

## 4.2 国外云计算商业发展状况

首先，是亚马逊。毫无疑问的，亚马逊在云计算领域具有着碾压性的优势。在亚马逊发布的 2017 年第二季度财报中，其 AWS 云计算部门斩获了 41 亿美元营收，同比增长了



42%，其中运营利润更是达到了 9.16 亿美元，同比增长了 28%。对此，亚马逊首席财务官布莱恩·奥萨维斯基表示，今年 AWS 部门的总营收有望达到 160 亿美元。这样的成绩在云计算厂商还没有集体实现盈利之前，无疑是难以令人望其项背的。正因为如此，亚马逊 CEO 杰夫·贝索斯一度问鼎世界首富的位置，而人们对于亚马逊股票的追捧，倒不如说是人们对于亚马逊 AWS 业务的追捧。在云计算领域，亚马逊 AWS 一直是 Gartner 魔力象限的领导者，并且远远超越其他云计算领域的竞争对手。

亚马逊 AWS 之所以能够长期占领云计算领域的制高点，原因包括早期的技术积累、丰富的细分市场、庞大的应用市场、对技术的敏感和坚持等等。而最重要的，应该是亚马逊的先发优势。在七八年前，这是一家随着 B2B 和 B2C 的浪潮而兴起的网上卖书和网上购物的公司，最初为了支撑其庞大的互联网网上购物业务，尤其是为了支持圣诞节的热销等具有庞大并发用户数量和访问的交易，亚马逊部署了大量冗余的 IT 计算和存储资源。后来他们发现这些资源在绝大部分时间里都是空闲的，为了充分利用闲置的 IT 资源，亚马逊建立起弹性云计算并对外提供效能计算和存储的租用服务，用户仅需要为自己所使用的计算平台的实际用量付费。这样的按需而定的付费，相比企业自己部署相应的 IT 硬件资源便宜很多。这就是以云计算基础设施作为服务（IaaS）的典型，是典型的因技术创新而带动的商业模式成功，至今仍是亚马逊最大的盈利来源之一。

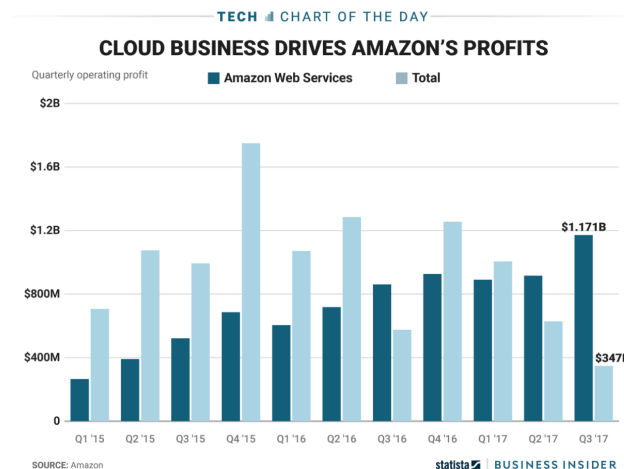


图9 亚马逊近年不同季度财报，可见其 AWS 业务的盈利增长

其次，是微软。在 IBM 遭遇营收连续 5 年下降的同时，微软第四财季营收为 247 亿美元，好于市场预期，其云平台 Azure 季度营收更是同比上涨 97%。不仅如此，微软建立在

Azure 之上的商业软件 Dynamics 365 和办公软件 Office 365 等商业云应用实现了 84 亿美元的季度营收，同比上涨 23%。以 Azure 和 Office 365 为代表的云服务在刚刚过去的财年，为微软贡献了 189 亿美元的收入，已经接近公司设定的 2018 财年商务云营收达到 200 亿美元的目标。几乎所有的人都认为，微软赌对了云计算。

微软在云计算方面的优势有以下几点：1. 企业级品质和面向企业客户：从数据中心建设，全球光纤网络到虚拟化技术（Hyper-V）到操作系统（Windows Server），再到运行云服务的各项基础软件（SQL, IIS），这些全部是微软自己研发的成熟的产品，都有几年甚至几十年在企业环境中运行的历史，其稳定性和可靠性是毋庸置疑的。2. 公有云和私有云混搭（即混合云）：微软的新产品 Azure 混合云，类似将现成的 Azure 私有云提供给用户，用户可以按照企业需要自己管理和使用，并且这个私有云，无论从接口和体验上，都与 Azure 高度统一，可以向公有云无缝扩展。这对很多有相关需求的大企业来说具有很强的吸引力。3. 与微软固有产品线的无缝整合。目前几乎所有的微软产品线都在考虑跟云沾上边，几乎所有的产品都实现了与 Azure 的整合，有些产品甚至直接转移到 Azure 上变成 SaaS/PaaS 云服务，比如 AD 转变为 AAD，SQL 转变为 SQL Azure 等等。所以 Azure 很快就形成了一个从 IaaS（VM, Network）到 PaaS（Storage, AAD, SQL, Media, VSOnline）到 SaaS（Office 365, Machine Learning）的一套极为完整的云生态体系。这些产品本来就有极大的用户群体，现在这些用户都可以轻松地迁移到 Azure 上，这就是 Azure 这几年都在飞速增长的重要原因。这种构建生态系统的商业模式，成为微软在云计算领域的强有力支柱。

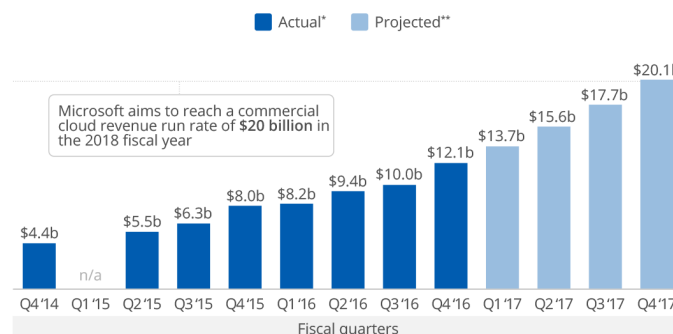


图 10 微软云服务近年各季度财报

然后，是谷歌。谷歌的云计算平台支持很强的容灾性，支持应用的快速自动部署和任务调度，能够提升多并发用户的高性能感受，而最关键的是，谷歌做到了每用户访问最低的运营成本。云计算使得 Google 的成本是其竞争对手的 1/40。这就是利用互联网服务吸取人气，采用后向广告收费的商业模式。谷歌用云计算平台构造了世界上最大的“超级计算机”，不仅便宜而且性能高，并且很难被复制，从而逐渐发展成为 PaaS 的商业模式。

不过，在云计算方面，谷歌更像是一个使用者而不是提供者。与亚马逊相比，谷歌在云计算业务方面已经有很大差距。Synergy Research 发现，谷歌在全球云基础设施中排名第四位，市场份额占 5%，落后于 AWS、Azure 和 IBM。由于目前规模较小，谷歌云计算业务还有很大的增长空间，目前正在包括印度和巴西在内的国家开放云服务。

谷歌在云计算方面是稍显落后的，究其原因，首先，谷歌并不是一个以服务见长的公司，这就与亚马逊产生了差距；其次，谷歌对云计算的重视程度远不及亚马逊；最后，谷歌的企业业务发展的一直不是很好，因为这一部分对谷歌来说，产出上不管是效率还是占比都很低，所以谷歌对企业市场的态度不如亚马逊和微软。

除了以上厂商，国外云计算市场中具有较强竞争力的企业还有 IBM、甲骨文、VMware 等。随着云计算的盈利能力的提升，各大企业在这一领域的竞争将会更加激烈。

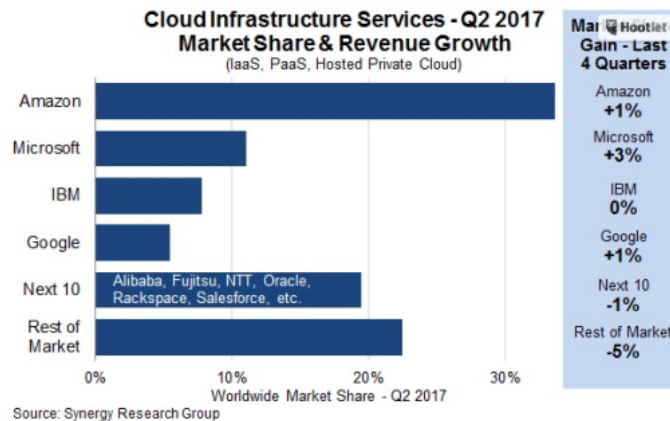


图 11 2017 Q2 各大云计算服务提供企业市场份额。亚马逊远超其他企业

#### 4.3 国内云计算商业发展状况

国内市场也有亚马逊、微软等企业的参与，这里只讨论国内企业。

国内云计算市场上，阿里云占据了半壁江山：国内云计算厂家并不算多，却在急速增长，但基本上是一超多强的局面：阿里云占据了超过 50% 的市场，腾讯云、金山云、华为

云等同属第二梯队，距离阿里云有较大的差距。阿里云目前在弹性计算领域的积累已经可比肩国际巨头 AWS、Azure，并且在国外市场也得到了认可。

与亚马逊类似，阿里是国内最早投身于云计算的企业。在 AWS 刚刚起步，远远没有如今的盈利能力时，阿里便开始了自己在云计算领域的投入。阿里精准的战略眼光带来的是在云计算领域的先发优势和马太效应。国内云计算领域的后起之秀很多，却都远远没有达到阿里云的高度，哪怕是国外企业算在内。2017 年 5 月 18 日晚间，阿里巴巴集团公布 2017 财年全年财报。其中，阿里云该财年营收规模达到 66.63 亿元人民币，同比增长 121%，连续两年实现三位数增长。截至 2017 年 3 月底，阿里云云计算付费用户数量达到 87.4 万，推动 2017 财年阿里云营收达到 66.63 亿元。自 2016 财年第一季度以来，阿里云已经连续 8 个季度保持着超过 100% 的高速增长，2017 财年第四季度的单季度营收为 21.63 亿元，接近阿里云上一个财年的整体营收 30.19 亿元。阿里云总裁胡晓明表示，阿里云现阶段的首要目标是保持对新技术的投资，并将阿里在亚洲市场的领先优势扩展至全球。



图 12 阿里云近年财报

技术层面上，在双十一、双十二等举世罕见的高并发、大流量场景下，淘宝等软件仍能正常运作，这说明了阿里云的可靠和技术优势，而这是很难被超越的。

尽管很多企业加大了对云计算领域的投入，也有很多新的云计算企业诞生，但阿里云在中国云计算市场上仍然是一家独大，并且领先第二很多。不知道这对于未来中国的云计算市场来说是福还是祸。

## 5. 未来展望

从当前的发展趋势和市场需求来看，云计算未来的发展趋势包括对移动领域的覆盖、传统企业基于云的转型、更大规模的云基础设施建设、数据安全问题的解决以及服务内容的革新等等。毫无疑问的是，云计算在未来会更加深刻的改变我们的生活。真正的“云时代”，即将到来。

## 6. 结论

本次调研报告，调研了当前云计算领域的相关情况。首先，介绍了云计算领域的基本情况，对云计算行业做了基本概述。然后，根据云计算的分类，对云计算及相关产品进行了更加详细和深入的介绍。接着，调研了云计算的相关关键技术，对云计算有了更深入的了解。然后，调查了国内外云计算市场的商业模式、市场份额、产业链和相关龙头企业情况。最后，对云计算未来的发展趋势进行了展望。

经过本次调研，对云计算的基本情况有了比较详尽的了解，既从技术层面上了解了云计算，又对当前国内外云计算市场的商业模式和企业情况有了比较深入的了解。本次的调研比较成功。

## 7. 参考资料

- [1] 陈康, 郑纬民. 云计算:系统实例与研究现状[J]. 软件学报, 2009, 20(05):1337-1348.
- [2] 张建勋, 古志民, 郑超. 云计算研究进展综述[J]. 计算机应用研究, 2010, 27(02):429-433.
- [3] 陈全, 邓倩妮. 云计算及其关键技术[J]. 计算机应用, 2009, 29(09):2562-2567.
- [4] 王意洁, 孙伟东, 周松, 裴晓强, 李小勇. 云计算环境下的分布存储关键技术[J]. 软件学报, 2012, 23(04):962-986.
- [5] 李乔, 郑啸. 云计算研究现状综述[J]. 计算机科学, 2011, 38(04):32-37.
- [6] 王佳隽, 吕智慧, 吴杰, 钟亦平. 云计算技术发展分析及其应用探讨[J]. 计算机工程与设计, 2010, 31(20):4404-4409.

- [7] 邱佳烁. 浅谈虚拟化技术与云计算的结合[J]. 产业与科技论坛, 2017, 16(23):58-59.
- [8] 钱童心. 亚马逊、微软、谷歌角逐云业务[N]. 第一财经日报, 2017-11-02(A07).
- [9] 刘晓燕. 企业雄心:亚马逊——世界上最非同寻常的公司——才刚刚起步[J]. 英语文摘, 2017, (06):24-28.
- [10] 本刊编辑部. 阿里云眼中的产业互联网发展趋势[J]. 信息化建设, 2017, (09):16-17.
- [11] 管理信息系统 课件