

云计算技术简述

戴元顺

美国田纳西大学

摘要 云计算(Cloud computing)是当今IT界的热门技术,借助云计算,网络服务提供者可以在瞬息之间,处理数以千万计甚至亿计的信息,实现和超级计算机同样强大的效能。同时,用户可以按需弹性地使用这些资源和服务,从而实现将计算作为一种公用设施来提供的梦想。本文首先介绍了云计算的基本概念、组成部分、技术要点和国内外发展情况,然后分别从服务器虚拟化、存储虚拟化、应用虚拟化、平台虚拟化和桌面虚拟化五个方面重点介绍虚拟化技术同云计算之间密不可分的关系。

关键词 云计算;虚拟化;云存储;服务计算

1 云计算概念简介

对于云计算,人们有着不同的理解和定义。本文认为,云计算是一种大规模资源整合的思想^[1-3]。云计算是IT界未来发展的必然趋势,所谓“天下大势,分久必合,合久必分”,IT界亦然^[4]。

1946年2月15日,世界上第一台电脑ENIAC诞生,占地十几间屋子,计算机行业也随着这个庞然大物应运而生。之后IT界开始了第一次“分”的过程,即个人电脑的诞生。随着网络的出现,IT界出现了第一次“合”的过程,如Client/Server架构、万维网WWW、Email等都是由高性能服务器通过网络为多个客户提供服务。但是,随着用户数量不断增加,服务器端的负荷越来越重,这种模式的瓶颈由于那个年代有限的资源(如带宽、计算、存储等)而变的越发严重,所以第二次“分”的时代来临,如分布式计算、点对点P2P技术、网格计算(Grid Computing)、Web2.0等。每个用户(Peer)既是被服务者同时也是服务者,大家共同分担庞大的计算、传输及存储任务。由此而诞生的企业有PPStream、YouTube、维基百科等。

为方便理解以上云的思想,可从图1的类比来进行阐述和表达。“服务器群”类似于“发电机”提供“电力”资源;虚拟技术类似于“变压装置”使电压成倍增加或降低,从而实现弹性计算;资源调度器类似于

“整流装置”,可以整合各个“发电站”的电力进行集中供电;服务管理器传送云服务,类似于“电力传输设备”,可以统一提供和管理IaaS(基础设施即服务)、PaaS(平台即服务)和SaaS(软件即服务)等一系列云服务;安全监控系统类似于“保险装置”,可以保证传输过来的“电”安全可靠,不会由于异常情况(如短路)损害家电和人身安全;云电脑、云手机等终端设备类似于家电,可以通过他们获取“电”(云资源)。

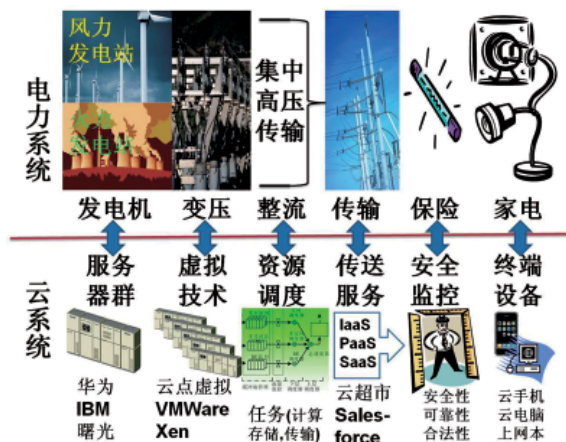


图1 云计算类比于电力系统

2 云计算在国内外发展情况简述

尽管云计算的思想已经孕育很久,但在美国取得蓬勃发展也只是最近五六年的事情。中国近一两年也在云计算方面有了长足发展。

2.1 云计算在美国的发展情况

2003年,美国国家科学基金(NSF)投资830万美元支持由美国七所顶尖院校提出的“网格虚拟化和云计算VGrADS”项目,由此正式启动了云计算的研发工作。基于此项目的后续开源产品有现在著名的云计算开源软件Eucalyptus(此软件在美国已经被广泛应用),美国航空航天局(NASA)的云计算系统Nebula^[5]。亚马逊公司的云计算也采用了相同技术(其客户端软件EC2可以直接连入Eucalyptus的服务器)。本文作者继续深入研究,相继推出了云点(Cloudian)系列产品(www.cloudian.org),涵盖了建设云平台所必须的各个模块,包括虚拟化、管理、调度、服务(IaaS、PaaS、SaaS)、安全、可信、云电脑和手机终端等。

从2004年开始,Amazon陆续推出了简单队列服务、MechanicalTurk等云计算服务雏形。云计算服务成熟的标志是亚马逊在2006年推出的简单存储服务(S3)和弹性计算云(EC2)。Google以应用托管、企业搜索以及其他更多形式向企业开放了他们的“云”。Google以发表学术论文的形式公开其云计算三大法宝:GFS、MapReduce和BigTable,并在美国、中国等高校开设云计算编程课程。目前,Google已经允许第三方在Google云计算中通过Google App Engine运行大型并行应用程序。2009年4月,谷歌推出了Google应用软件引擎(Google AppEngine,下称GAE),这种服务让开发人员可以编译基于Python的应用程序。Microsoft推出了Windows Azure操作系统,这个系统作为微软云计算计划的server端操作系统(Cloud OS)为广大开发者提供服务。同时,微软也推出了免费在线手机同步备份业务MyPhone。微软为使其在互联网上与Google竞争,MS宣布推出数据存储及网络管理软件Live Mesh,迈出从PC领域到云计算的一大步。Microsoft将Live Mesh视为基于网上数据中心的软件平台,可以提供多种服务,包括计算机远程控制、电子设备及数据存储等。IBM于2008年提出了“蓝云”计划,推出共有云和私有云的概念。IBM提出私有云解决方案是为减少诸如数据、信息安全等共有云现存问题,从而抢占企业云计

算市场。将重心放在将现有产品和技术整合上,并充分利用开源产品,是IBM 云计算方案的特点。例如利用Xen虚拟化产品、Linux操作系统+X86架构组成的IBM 服务器云数据中心。在共有云方面,IBM 于2009年发布了LotusLive Engage。基于云端的协作平台是IBM 2010年发展的一个重点。Apple是云计算领域的另一位角色,从近年来推出的iTunes服务,到MobileMe服务,到收购在线音乐服务商Lala,再到最近在美国北卡莱罗纳州投资10亿美元建立新数据中心的计划,无不显示其进军云计算领域的巨大决心。

2.2 云计算国内运营商发展情况

与美国相比,国内的云计算发展虽处于起步阶段,但各大通信运营商都表现得异常活跃。中国移动推出了“大云”(Big Cloud)云计算基础服务平台,中国电信推出了“e云”云计算平台,中国联通则是推出了“互联云”平台。

2009年8月,中国移动研究院发布了0.5版本“大云”系统,并在此基础上设计了很多管理软件和应用。中国移动的“大云”是基于一些重要的开源软件开发的。以此为基础,中国移动研发和试验了高效的云系统管理软件和比较重要的几个应用,比如并行数据挖掘、云存储、大容量数据库及对搜索引擎的研究。中国移动通信研究院通过深入研发和试验HyperDFS、MapReduce、HugeTable、CloudMaster等云计算平台关键技术,自主搭建了1000个CPU组成的256节点规模的云计算试验平台,并开展系统评估与优化,构建了基于云计算技术的移动互联网业务海量数据存储和处理试验平台,开展了云计算应用研发和试验。

2009年9月22日,中国电信上海分公司携手EMC推出“e云”。这片“云”能按照用户的设定,自动利用电脑空闲时间,将信息备份到上海电信的“e云”数据中心,当用户遇到电脑破坏、数据破坏、误删除、在家办公、远程办公等情况时,只要通过网络连接至电信服务器,就可以在任何地方恢复任意一个时间点的数据。

“e云”可以把天翼Live等IM工具连接起来,目前,针对主流手机终端以及其他增值服务的开发都在进行中,

到时用户可以实现单一账号式登入。这意味着,未来不论是联通用户还是移动用户,无论是通过EVDO还是TD或WCDMA方式传输,只要设置一个单一账号,都能接入和使用电信的云服务。中国联通研究院也开展了“互联云”的试验。

3 云计算和虚拟化

虚拟化(Virtualization)技术是云计算系统的核心组成部分之一,是将各种计算及存储资源充分整合和高效利用的关键技术。

虚拟化的定义:虚拟化是为某些对象创造的虚拟(相对于真实)版本,比如操作系统、计算机系统、存储设备和网络资源等。它是表示计算机资源的抽象方法,通过虚拟化可以用与访问抽象前资源一致的方法访问抽象后的资源,可以为一组类似资源提供一个通用的抽象接口集,从而隐藏属性和操作之间的差异,并允许通过一种通用的方式来查看和维护资源。

3.1 服务器虚拟化

服务器虚拟化技术可以使一个物理服务器虚拟成若干个服务器使用,如图2所示。服务器虚拟化是基础设施即服务(Infrastructure as a Service,IaaS)的基础。

服务器虚拟化需要具备以下功能和技术:

1) 多实例:在一个物理服务器上可以运行多个虚拟服务器。

2) 隔离性:在多实例的服务器虚拟化中,一个虚拟机与其他虚拟机完全隔离,以保证良好的可靠性^[6-7]及安全性^[8]。

3) CPU虚拟化:把物理CPU抽象成虚拟CPU,无论任何时间一个物理CPU只能运行一个虚拟CPU的指令。而多个虚拟机同时提供服务将会大大提高物理CPU的利用率。

4) 内存虚拟化:统一管理物理内存,将其包装成多个虚拟的物理内存分别供给若干个虚拟机使用,使得每个虚拟机拥有各自独立的内存空间,互不干扰。

5) 设备与I/O虚拟化:统一管理物理机的真实设备,将其包装成多个虚拟设备给若干个虚拟机使用,响应每个虚拟机的设备访问请求和I/O请求。

6) 无知觉故障恢复:运用虚拟机之间的快速热迁移技术(Live Migration),可以使一个故障虚拟机上的用户在没有明显感觉的情况下迅速转移到另一个新开的正常虚拟机上。

7) 负载均衡:利用调度和分配技术,平衡各个虚拟机和物理机之间的利用率。

8) 统一管理:由多个物理服务器支持的多个虚拟机的动态实时生成、启动、停止、迁移、调度、负荷、监控等应当有一个方便易用的统一管理界面。

9) 快速部署:整个系统要有一套快速部署机制,对多个虚拟机及上面的不同操作系统和应用进行高效部

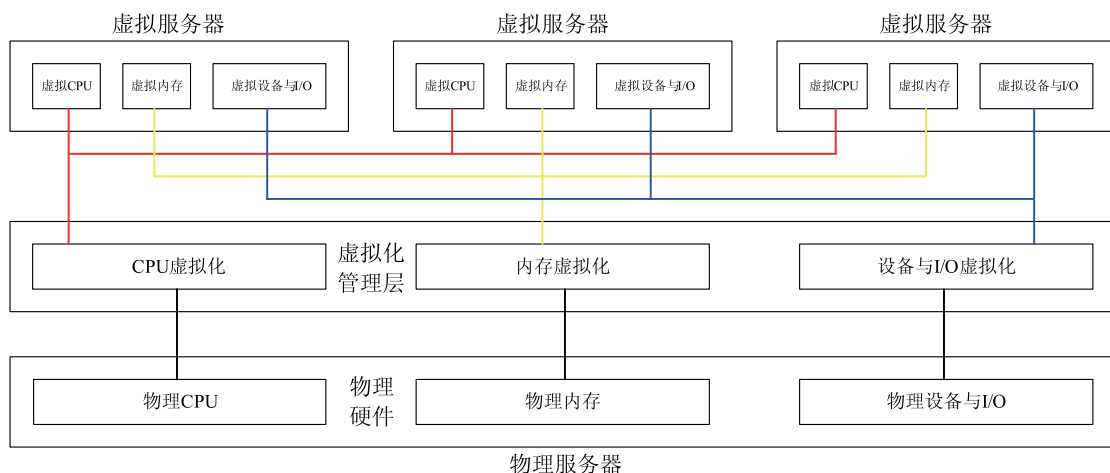


图2 服务器虚拟化

署、更新和升级。

3.2 存储虚拟化

存储虚拟化的方式是将整个云系统的存储资源进行统一整合管理，为用户提供一个统一的存储空间，如图3所示。

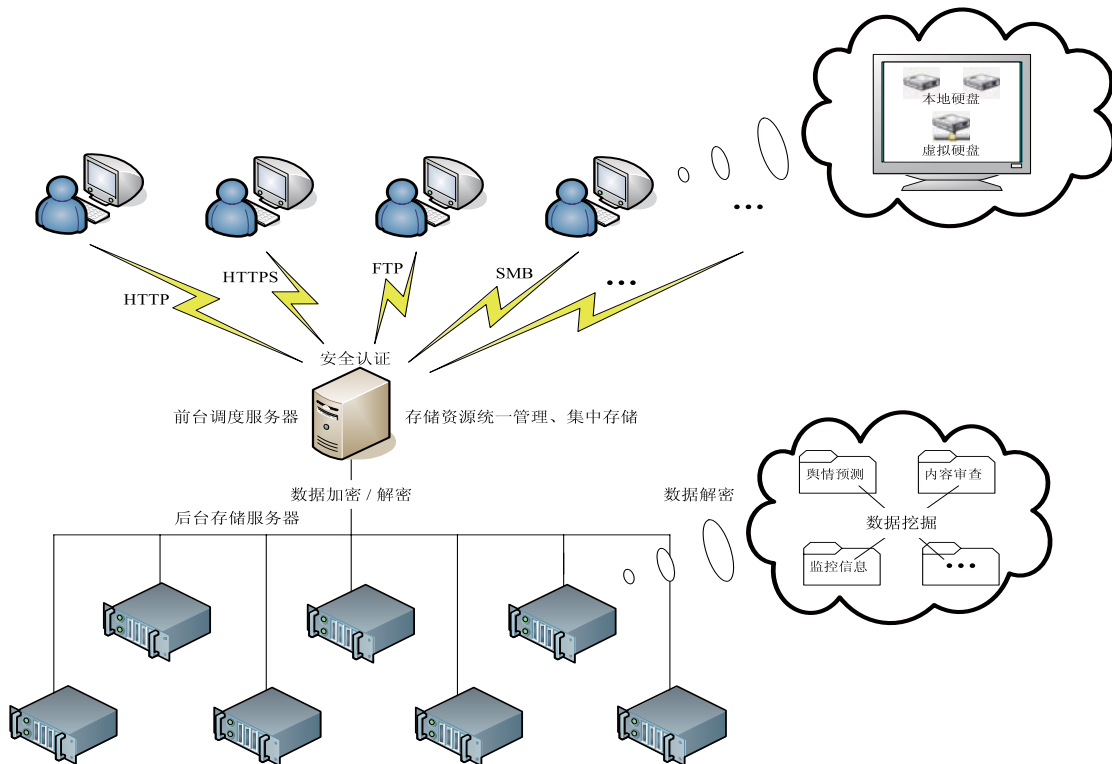


图3 存储虚拟化

存储虚拟化具有以下功能和特点：

1) 集中存储：存储资源统一整合管理，集中存储，形成数据中心模式。

2) 分布式扩展：存储介质易于扩展，由多个异构存储服务器实现分布式存储，以统一模式访问虚拟化后的用户接口。

3) 节能减排：服务器和硬盘的耗电量巨大，为提供全时段数据访问，存储服务器及硬盘不可以停机。但为了节能减排，需要利用更合理的协议和存储模式，尽可能减少开启服务器和硬盘的次数。

4) 虚拟本地硬盘：存储虚拟化应当便于用户使用，最方便的形式是将云存储系统虚拟成用户本地硬盘，使用方法与本地硬盘相同。

5) 安全认证：新建用户加入云存储系统前，必须

经过安全认证并获得证书。

6) 数据加密：为保证用户数据的私密性，将数据存到云存储系统时必须加密。加密后的数据除被授权的特殊用户，其他人一概无法解密。

7) 级层管理：支持级层管理模式，即上级可以监

控下级的存储数据，而下级无法查看上级或平级的数据。

3.3 应用虚拟化

应用虚拟化是把应用对底层系统和硬件的依赖抽象出来，从而解除应用与操作系统和硬件的耦合关系。应用程序运行在本地应用虚拟化环境中时，这个环境为应用程序屏蔽了底层可能与其他应用产生冲突的内容。应用虚拟化是SaaS的基础。应用虚拟化需要具备以下功能和特点：

1) 解耦合：利用屏蔽底层异构性的技术解除虚拟应用与操作系统和硬件的耦合关系。

2) 共享性：应用虚拟化可以使一个真实应用运行在任何共享的计算资源上。

3) 虚拟环境：应用虚拟化为应用程序提供了一个

虚拟的运行环境，不仅拥有应用程序的可执行文件，还包括所需的运行环境。

4) 兼容性：虚拟应用应屏蔽底层可能与其他应用产生冲突的内容，从而使其具有良好的兼容性。

5) 快速升级更新：真实应用可以快速升级更新，通过流的方式将相对应的虚拟应用及环境快速发布到客户端。

6) 用户自定义：用户可以选择自己喜欢的虚拟应用的特点以及所支持的虚拟环境。

3.4 平台虚拟化

平台虚拟化是集成各种开发资源虚拟出的一个面向开发人员的统一接口，软件开发人员可以方便地在这个虚拟平台中开发各种应用并嵌入到云计算系统中，使其成为新的云服务供用户使用，如图4所示。

定的测试环境中经过完整的测试才行。

4) 服务计费：完整合理的计费系统可以保证服务提供者获得准确的收入，而虚拟平台也可以得到一定比例的管理费。

5) 排名打分：有一整套完整合理的打分机制对各种服务进行排名打分。排名需要给用户客观的指导性意见，严禁有误导用户的行为。

6) 升级更新：允许服务提供者不断完善自己的服务，平台要提供完善的升级更新机制。

7) 管理监控：整个平台需要有一个完善的管理监控体系以防出现非法行为。

3.5 桌面虚拟化

桌面虚拟化将用户的桌面环境与其使用的终端设备解耦。服务器上存放的是每个用户的完整桌面环境。用

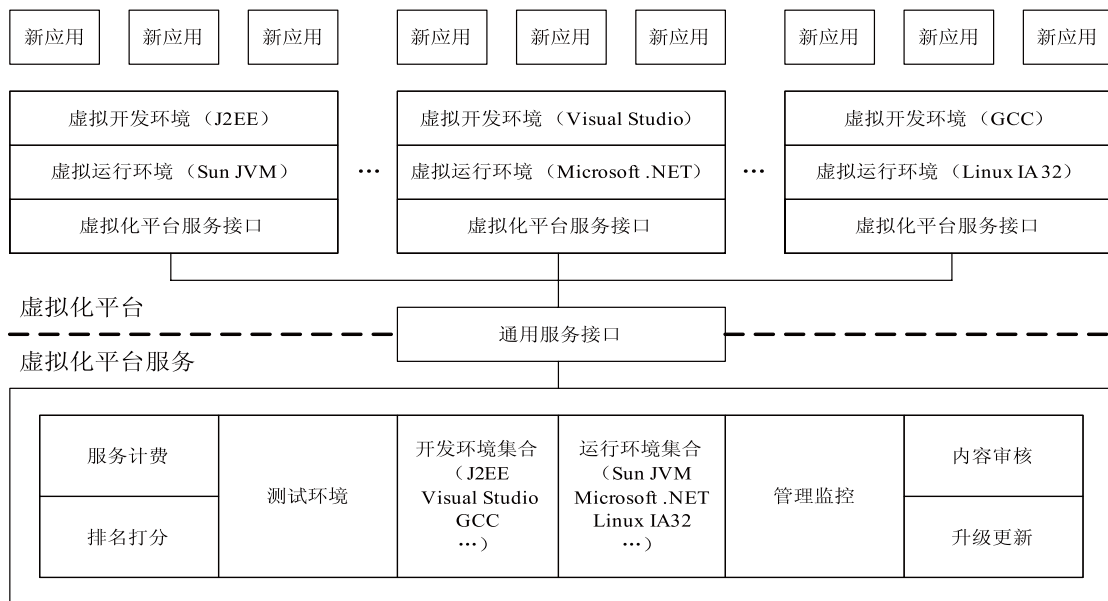


图4 平台虚拟化

平台虚拟化具备以下功能和特点：

1) 通用接口：支持各种通用的开发工具和由其开发的软件，包括C、C++、Java、C#、Delphi、Basic等。

2) 内容审核：各种开发软件(服务)在接入平台前都将被严格审核，包括上传人的身份认证，以保证软件及服务非盗版、无病毒及合法性。

3) 测试环境：一项服务在正式推出之前必须在一

户可以使用具有足够处理和显示功能的不同终端设备通过网络访问该桌面环境，如图5所示。

桌面虚拟化具有如下功能和接入标准：

1) 集中管理维护：集中在服务器端管理和配置PC环境及其他客户端需要的软件可以对企业数据、应用和系统进行集中管理、维护和控制，以减少现场支持工作量。

2) 使用连续性：确保终端用户下次在另一个虚拟

机上登录时,依然可以继续以前的配置和存储文件内容,让使用具有连续性。

施(硬件、平台、软件)以及建立在基础设施上的信息服务。提供各类资源的网络被称为“云”, 在使用者看

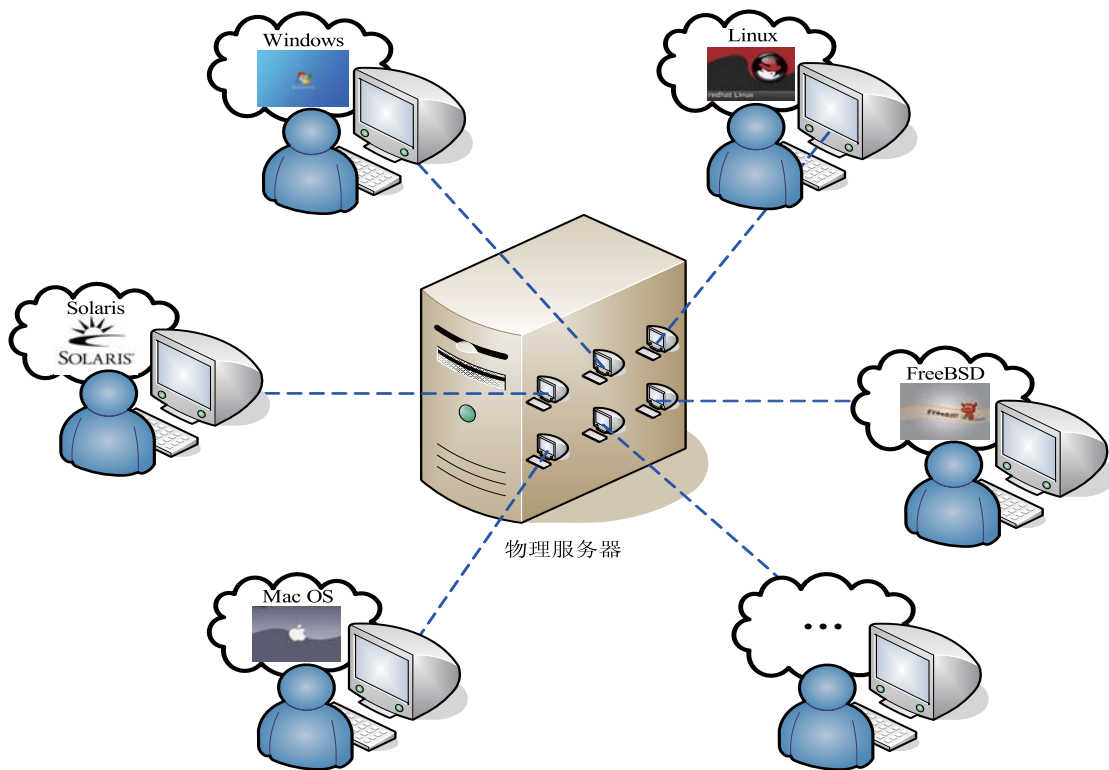


图5 桌面虚拟化

3) 故障恢复:桌面虚拟化是用户的桌面环境被保存为一个个虚拟机,通过对虚拟机进行快照和备份,就可以快速恢复用户的故障桌面,并实时迁移到另一个虚拟机上继续进行工作。

4) 用户自定义:用户可以选择自己喜欢的桌面操作系统、显示风格、默认环境,以及其他各种自定义功能。

本质上讲云计算带来的是虚拟化服务。从虚拟化到云计算的过程,实现了跨系统的资源动态调度,将大量的计算资源组成IT资源池,用于动态创建高度虚拟化的资源供用户使用,从而最终实现应用、数据和IT资源以服务的方式通过网络提供给用户,以前所未有的速度和更加弹性的模式完成任务。

4 总结

云计算是一种思想,一种大规模资源整合的思想,是IT界未来发展的必然趋势。云计算包括信息基础设

来,“云”中的资源是可以无限扩展的,并且可以随时获取、按需使用、弹性扩展和按使用付费。“云”服务就好比单台发电机模式转向电网集中供电的模式,它意味着计算能力也可以作为一种商品进行流通,就像煤气、水电一样,取用方便,费用低廉。云计算是并行计算、分布式计算和网格计算的发展,或是这些计算科学概念的商业实现。云计算也是虚拟化、效用计算、服务计算等概念混合演进并跃升的结果。

参考文献

- [1] Dai Y S,Xiang Y P,Zhang G W.Self-Healing and Hybrid Diagnosis in Cloud Computing[J].Lecture Notes of Computer Science,2009,5931:45-56
- [2] Yang B,Tan F,Dai Y S,et al.Performance Evaluation of Cloud Service Considering Fault Recovery[J].Lecture Notes of Computer



- Science, 2009, 5931: 571–576
- [3] Li H, Dai Y S, Tian L, et al. Identity-Based Authentication for Cloud Computing[J]. Lecture Notes of Computer Science, 2009, 5931: 157–166
- [4] Zou X, Dai Y S, Pan Y. Trust and Security in Collaborative Computing[M]. Hackensack, NJ, USA.: World Scientific, 2008, ISBN: 981–270–368–3
- [5] 美国航空航天局(NASA)的云计算系统(Nebula) [EB/OL]. [2010–02–25]. <http://nebula.nasa.gov/blog/2009/nov/how-eucalyptus-enables-ec2-compatibility-with-nebu/>
- [6] Dai Y S, Pan Y, Raje R. Advanced Parallel and Distributed Computing: Evaluation, Improvement and Practices[M]. New York: Nova Science Publishers, 2007
- [7] Xie M, Dai Y S, Poh K L. Computing Systems Reliability: Models and Analysis[J]. New York: Springer, 2004
- [8] Dai Y S, Levitin G, Trivedi K S. Performance and Reliability of Tree-Structured Grid Services Considering Data Dependence and Failure Correlation[J]. IEEE Transactions on Computers, 2007, 56(7): 925–936

作者简介

戴元顺

2000年获清华大学本科学位；2003年获新加坡国立大学博士学位；2004年7月至2007年5月，任美国普渡大学助理教授，兼美国电子信任和网络安全实验室(TEGO)主任；2007年6月至今，任美国田纳西大学助理教授，并在美国国家级重点实验室(ICL)建立了自己的研究团队；2009年5月至今，任电子科技大学讲座教授，协同自主计算实验室主任；现担任《IEEE Transactions on Reliability》副主编及多个国际期刊编辑或编委，曾发表学术论文70余篇，出版专著5部。

The Brief Review of Cloud Computing Technologies

Dai Yuanshun | The University of Tennessee, USA

Abstract Cloud computing is an emerging technology and of great concern by the whole IT world today. By cloud computing, the providers of network services can efficiently handle thousands of or even millions of tasks with the same power as a supercomputer. Meanwhile, users can utilize these resources and services in an elastic and flexible manner, so that to realize the dream of IaaS (Infrastructure as a Service) provision. This paper introduces the concept, components, detailed technologies and state-of-the-art in the field of cloud computing; and especially, it elaborates the close relationship between the technology of virtualization and the cloud computing.

Keywords Cloud Computing; Virtualization; Cloud Storage; Service Computing