

数据云技术发展的若干关键问题

匡松 周启海 刘小麟 徐畅畅 刘莹 赵华生

(西南财经大学经济信息工程学院 成都 610074)

摘要 数据云技术建立在传统存储和互联网存储的基础之上,并注入无限量概念,以期成为全球网络用户的后台内存与硬盘,它将使每个人都能够通过互联网迅速链接和同步共享海量数据。介绍了数据云及其未来发展趋势,重点讨论了影响数据云技术发展的若干关键问题。

关键词 数据云,互联网存储,网络硬盘,无限量

中图分类号 TP393.4 **文献标识码** A

Some Key Problems about Development of Data Cloud Technology

KUANG Song ZHOU Qi-hai LIU Xiao-lin XU Chang-chang LIU Ying ZHAO Hua-sheng

(School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 610074, China)

Abstract The data cloud technology is established on the basis of traditional memory and the Internet memory, and adds in the concept of infinitude, in order to be taken as the global network users' backstage memory and hard disk, it will enable each people rapidly link and synchronistically share mass data through the Internet. The data cloud and its future development trend were introduced, and some key problems which would affect the development of data cloud technology were discussed as the focal points.

Keywords Data cloud, Internet memory, Network hard disk, Infinitude

1 数据云技术概述

2007年,数据云技术(Data Cloud)作为十大科技新概念之一被提出。信息工业观察家 Thomas Vander Wal 于2006年首次提出数据云的原形——个人信息云(personal info-cloud)^[1],这种技术将个人信息分散地存储在互联网服务器上,并在用户需要时将所需数据同步重构。从根本上讲,不管是个人信息云,还是应运而生的数据云,其提出和发展都是建立在存储技术之上。纵观存储技术的发展历史,到目前为止,已经经历了以下两个重要的阶段:(1)传统的 DAS(直接连接存储, Direct Attached Storage)方式完全以服务器为中心,通过线缆直接连接到计算机处理器,并寄生在服务器或客户端上。随着应用系统的扩大和数据量的增长,这种实质上由硬盘堆叠起来的存储方式,在用户数量增加或不间断提供服务时,充分暴露了响应缓慢的缺陷。(2)随着用户对数据服务的要求不断提高,存储方式由传统直连式存储走向分离式存储的趋势日益明显。快速发展的网络时代催生了以数据和网络为中心的存储技术——网络附加存储 NAS(Network Attached Storage)和存储区域网络 SAN(Storage Area Network),如图1所示。

NAS也称网络磁盘阵列,它是一种专业的网络文件存储及备份设备。图1表明,NAS将存储设备通过标准的网络拓扑结构连接,无需服务器便可直接上网。因此,它使整个系统

的管理和设置较为简单,可以有效地实现多用户、多应用的单一数据的共享,并可以在 Unix 和 Windows 客户机上使用,实现异构机(包括其软硬件)的存储、访问与共享。

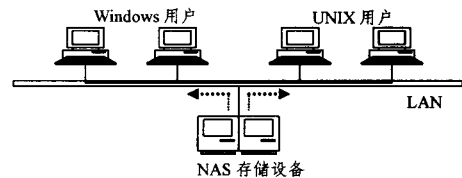


图1 NAS 结构示意图

以数据存储为中心的 SAN 是存储设备相互连接且与一台服务器或一个服务器群相连的网络,它采用可伸缩的网络拓扑结构,如图2所示。通过具有高传输速率的光纤通道直接连接网络服务器和诸如大磁盘阵列或备份磁带库的存储设备^[2]。

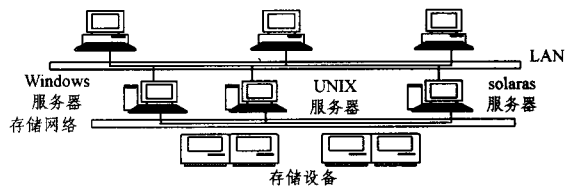


图2 SAN 结构示意图

到稿日期:2008-09-28

匡松(1957-),男,教授,硕士生导师,主要研究方向为计算机应用,E-mail:ksong@swufe.edu.cn;周启海(1947-),男,教授,博(硕)士生导师,主要研究方向为财经计算、算法研究与应用、同构化信息处理、计算几何等。

可以预测未来的数据云将是继这两次存储技术之后又一轮存储技术创新浪潮。尽管迄今为止,关于数据云业内仍然没有形成公认的定义和共识,但其定义基本特色与关键点仍可概括为:数据云是由可以通过互联网设备进行即时交换的、可随时随地使用的数据组成的无限量数据集。也就是说,数据云技术至少是建立在目前存在的两种互联网存储基础上又对其传输、存储和计算等性能加以改进的网络存储技术。而数据云技术的核心,是在现有互联网存储技术的基础上实现5个“无限量”——CPU处理器无限量、硬盘无限量、传输信道无限量、内存无限量和用户无限量。

2 数据云技术的发展雏形——互联网存储

目前,微软、Google、EMC等众多知名公司都已对数据云技术开展研究,但数据云技术在全球范围内仍处于研发阶段,还没有成形的数据云产品。人们也不能完全想象出这种技术的未来轮廓,似乎我们现在距离真正的数据云还很远。但是,对于数据云的雏形产品——互联网存储,却早已为大众所熟悉,并且已经在国内外发展得如火如荼。

2.1 中国的网络硬盘发展情况

随着硬盘价格的缩水和网络存储的兴起,一个个提供网络硬盘服务的网站如雨后春笋般涌现出来。目前,我国比较有特色的网络硬盘产品主要有以下几种:

(1)中国活动通信 Mofile

Mofile是国内最早的网络硬盘,已服务于全球数千万用户。因为Mofile提供的免费永久有效的文件储存空间只有64M,而其临时存放空间最大达到1G,速度也比较快,所以Mofile是一个不错的文件中转站。2006年,以网络存储闻名于世的Mofile又开始了播客网站的经营。

(2)存储在线 Dostor.com

“存储在线”是目前全球唯一一家专注于提供存储信息和服务的中文存储资讯平台,致力于让全球华人“把握存储未来”。“存储在线”旨在通过提供全面、及时和公正的存储产品和技术资源相关的中文信息,利用高效、方便的网络工具,建立信息多元化、使用简单快捷的中文网络空间,成为世界各地华人最好的存储网络平台。

(3)中容数据 ChinaVaultTM^[3]

2005年,Storage Networks的创始人Peter Bell和刘一男通过整合数据存储和数据信息生命周期管理的理念,创立了威泰科技的数据存储服务品牌——中容数据。中容数据改进了数据存储模式,将加密后的企业数据在线备份至中容数据保管中心和全球合作伙伴的数据中心,从而为广大用户提供尖端的数据保护服务,实现了最优化的数据保护和恢复能力。

2.2 国外的网络硬盘发展现状

互联网的普及让信息世界的国界定义更加淡化,相信以下介绍的国外著名的网络硬盘已经或者将为广大中国网民带来便利^[4]。

(1)MediaFire

世界最大的在线媒体中心Streamload旗下的MediaMax是一个口碑较好的在线存储与分享服务网站。一旦用户将自己计算机上的一个文件夹与MediaMax连到了一起,文件夹里的所有文件会在网络上自动实现后台同步存储,而用户在前台可以完全不受影响地正常工作。文件上传成功后Media-

Fire系统会发送一封Email到用户邮箱,注明上传文件的下载、删除地址等相关信息,便于用户之间分享资源。

(2)Xdrive

Xdrive是一个在网上免费而且安全存放文件的空间,可以想象Xdrive就是自己的个人硬盘,可以随时通过任何方式使用Xdrive,包括台式机、笔记本和使用无限通讯协议WAP的手机等。只需要从Xdrive.com网站上下载一个客户端软件,即可在资源管理器中创建一个X盘,通过简单的操作,可以将该网站的个人目录虚拟成PC机上的一个本地磁盘,并完全集成到资源管理器中,轻松地进行各种操作和管理。

(3)Box.net

在Box.net的主页上,醒目地挂着广告语:Work on documents anywhere,Protect your important files,Collaborate and share with anyone,体现着其共事、保护、合作、分享的主旨。Box.net是一个口碑较好的网络存储服务,为每个注册用户免费提供1GB高速稳定的在线存储空间,在测试期便已受到网络存储用户的追捧。但是从2006年起,注册Box.net已经再没有免费的午餐了。

(4)OmniDrive

除界面之外,OmniDrive毫不逊色于Box.net。因此,自从Box.net停止开放免费注册后,OmniDrive自然成为一款能与Box.net相媲美的替代品。OmniDrive的界面风格与XP的风格相似,提供3种文件查看方式,分别是图标方式、详情方式及缩略图方式,支持右键弹出菜单。OmniDrive的一大特色是通过包含RSS的网页链接来同步更新目录。实际上,OmniDrive是一款已经开发很久的产品,不过到目前为止还是没有完全对外正式开放。

(5)GDrive和Live Drive

虽然Google一直在推出各式各样的服务,但是已经很少有Gmail这样震惊市场的产品了。而传说中的GDrive有可能成为继Gmail之后的下一部狂潮力作。2006年3月,GDrive端倪初现,终于揭开了其神秘的冰山一角:这项服务代号为GDrive鸭嘴兽的产品,是一款可以通过互联网访问的存储驱动器,用户可以在该服务中存储大量各种类型的文件或者与好友共享。

不知是偶然,还是两大巨头的竞争,微软公司的实时硬盘Live Drive计划也正在开发两项在线存储服务的公共测试版;SkyDrive提供了公共、私有和共享文件夹的在线文件存储服务;而FolderShare提供了受控制的对等文件同步功能。

2008年2月,微软向雅虎提出收购之后仅两天,科技巨人Google马上积极采取对策,并计划在这场收购案中扮演破坏者的角色。两者的对立在一定程度上解释了它们同时研发互联网存储的“偶然事件”的必然性,似乎也预示着将来在数据云市场的另一番巨头之间的较量。

3 数据云的前景困难重重

虽然业内对于数据云技术的前景普遍看好,但是从如今互联网存储并不普及的现状到未来形成数据云笼罩全球的大好局面,这片“数据云”要走的路还很长,路上也一定会荆棘遍布。

3.1 成本

在科技领域,成本是每一项技术由研发到投入使用首先

应考虑的问题。到目前为止,已经有不计其数研发完成的新技术因其成本太高而被扼杀在摇篮里。这也是数据云所面临的现实问题。通俗地说,未来的数据云就是要取代如今计算机中的内存和硬盘间相互传输数据的功能,通过网络实现计算机与数据云之间的交换,完成数据存储和计算,并将结果返回给计算机用户。正如目前呼声越来越高的网格技术一样,在单个或几个企业内部使用网格计算会达到优化系统、简化管理、降低成本的目标,但是一旦其应用区域超过了企业范围甚至变为服务于全球性的技术,则对服务器的空间大小和响应速度提出了完全不同的要求,同时需要高速网络传输和数据安全保证,因此大范围应用该技术的边际收益很可能成为负数,整体成本也直冲而上。数据云技术也面临着相同的问题。这种规模效应的递减在经济学中是一个非常简单的道理,同时也应该引起数据云技术研究者的高度重视。否则,数据云将来也会面临有技术支持无市场需求的窘境。

3.2 异构性

据初步估算,仅中国国内就有近 60 家大大小小的硬件生产厂家,全球范围内的厂家数量更是大得惊人。目前,业内对于同类硬件商品并无统一生产标准,众多厂家生产的计算机硬件必然具有异构性。退一步讲,即使全球只有两家公司生产硬件,只要不能形成统一的制造标准,就会对数据云技术的应用产生巨大阻力。未来的数据云作为一个全球性的统一服务器和网络硬盘,很难有效分辨所有硬件的异构部分并加以区别处理,这是数据云在技术方面遇到的最大难题。例如,如果每一台计算机的声卡不同,所需驱动程序就会不一样,因而面对通过数据云传输而来的相同声音数据,必定会出现一部分“哑机”的情况。

3.3 安全问题

数据云的安全威胁主要来自两个方面:第一,互联网故障或中断造成的用户数据损坏或丢失;第二,由于计算机病毒、黑客软件等造成用户的秘密文件被窃取。

互联网发生故障的概率会随着互联网服务器的不断升级以及互联网传输技术的不断改进而降低。但是故障和中断不仅仅来源于技术方面,更多的还来自于自然灾害和人为破坏。谁又能保证将来为我们服务的数据云会像硬盘一样忠诚地始终随着计算机的启动而启动呢?

从网联网开始应用的第一天到现在,虽然各种防火墙、密码认证等安全保障技术层出不穷,但是“道高一尺,魔高一丈”,计算机病毒和黑客软件对个人数据的威胁有增无减,如果数据云的管理者不能向用户保证数据的完整性和安全性,用户肯定不能产生对它的完全信赖;反过来,即使有人敢保证用户把自己的东西存在网络上是完全安全可靠的,聪明或谨慎的用户就会相信吗?

3.4 时间考验

目前各大研究机构所憧憬的数据云是一种能够取代现有 PC 机硬盘、通过服务器与 PC 机之间的数据交换实现数据处理和存储的技术,这就意味着计算机硬盘将会完全被排斥于 PC 机之外。

目前全球平均互联网普及率仅为 19.1%,全球未连接互联网的 PC 机数量众多,要让互联网的光芒洒到地球的每个角落,尚需假以时日。因此,数据云技术目前只能在小范围内发展,而在全球范围内其发展所需要的网络环境暂时还不具备。数据云的研究者们必须耐得住寂寞,经得起时间的考验,等待网络环境成熟之时,再将数据云推出市场。

另一方面,数据云完全代替硬盘和 U 盘的道路也很艰难,硬盘、U 盘存储上的文件资料取用方便且更让用户放心,在这点上目前技术上定义的数据云还远远无法和其竞争。数据云技术必须有一种包容力和毅力,也许数据云与硬盘、U 盘共同掌管计算机存储世界的时间将会维持很长。

结束语 数据云技术正在成为存储领域的又一次巨大创新和革命。它在现有的限量网络存储基础上,将令每个用户都能够通过互联网迅速链接资源并且及时得到响应而且同步共享海量数据,而不受任何用户自身设备的制约。从这个角度来说,数据云要承担现有 PC 内存和硬盘的双重任务,既要作为无限网络内存,通过服务器实现用户数据的计算和存取;又要作为无限网络保险柜,存储用户需要长时间保存的数据并保障其安全。可以构想,尽管还面临着不少亟待攻关的难题,但数据云的未来大范围应用过程与方式,会通过现有的网络存储、加密保护、网络计算、直接内存存取等以及将来可能出现的网络交换技术的大整合,而将整个世界虚拟化并重构成一个超级巨型计算机群世界。这个超级巨型计算机群,拥有分布在全球各地的、高度灵活的前台操作和显示工具,也具有数据云这个无限量内存和硬盘作为强力支撑与高效处理的强大后台数据基地。

总之,随着数据云技术的日渐进步与日益成熟,它终将为 PC 机“减肥”和网络“健身”,使它们的内存和硬盘无限扩充,从而为人们带来无限量的共享资源,并将促使全球更快地迈向更高阶段的网络化世界。

参 考 文 献

(上接第 256 页)

[10] Kasin O, Scott J M. Scheduling Batch Processing Machines in Complex Job Shops[J]//Proceedings of the 2001 Winter Simulation Conference, 2001:1200-1207
[11] Oey k, Mason S J. Scheduling batch processing machines in complex job shops[J]//Proceeding of the 2001 Winter Simulation

[1] Roush W. The Internet Is Your Next Hard Drive, Technology Review [J], July 2006
[2] DAS, NAS, SAN 存储系统完全分析. <http://biz.chinabyte.com/128/2300128.shtml>
[3] 九陆. 中容数据保驾“在线存储”. 通信产业报, 2006-03-27
[4] Xdrive 网络硬盘[J]. 计算机采购, 2000(43)
Conference, 2001:1200-1207
[12] Jens Z, Lars M. Simulation-based selection of machine criticality measures for a shifting bottleneck heuristic[J]//Proceedings of the 2006 Winter Simulation Conference, 2006:1848-1854
[13] 黄文奇, 黄志. 作业车间调度问题的一种改进的转移瓶颈算法[J]. 计算机工程与应用, 2005(2):59-62