

云存储技术探讨

张龙立

(国脉通信规划设计有限公司 哈尔滨 150040)

摘要

本文基于云计算与云存储的概念,具体分析云存储系统构架模式、典型拓扑、技术优势及特点,并与传统的存储架构模式进行了对比,并就未来云存储发展趋势作了深入探讨。

关键词 云计算;云存储

1 引言

全球数据量的猛增使得数据存储日益成为各企业关注的专业问题,持续增长的数据存储压力带动了整个存储市场的快速发展。企业的存储设备越来越多,对管理和维护的投入也越来越大。为了保证数据安全和业务的连续性,我们需要建立相应的数据备份系统和容灾系统。除此之外,对存储设备进行定期的状态监控和维护,软、硬件的更新和升级也是必须的,这些都需要专业的技术人员,增加了企业数据维护、升级和管理费用。

2 云计算与云存储

云计算是分布式处理(distributed computing)、并行处理(parallel computing)和网格计算(grid computing)的发展,是透过网络将庞大的计算处理程序自动分拆成无数个较小的子程序,再交由多台服务器所组成的庞大系统经计算、分析后,将处理结果回传给用户。通过云计算技术,网络服务提供者可以在数秒之内,处理数以千万计甚至亿计的信

息,达到和“超级计算机”同样强大的网络服务功能。

云计算的设计实施方向是将在不同地域和情况下各种台式机、单一服务器及个人计算资源,通过虚拟化方式统一到庞大的计算资源“云”系统中来。云系统统一调配、协调、处理来自不同客户的运算需求,通过云系统计算后对外输出运算结果。从整体来看,云计算系统可以认为是以数据处理、数据运算为中心的系统。

云计算系统不但能对数据进行处理和运算,系统中还有大量的存储阵列设备,以实现计算数据的保存和管理。在云计算系统中配置相应的存储设备,该计算系统即拥有了云存储系统功能。由此可以理解,云存储是以数据存储和数据管理为中心的云计算系统。

云存储系统通过集群应用、网格技术或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应用软件集合起来协同工作,共同对外提供数据存储和业务访问功能。云存储是云计算系统的延伸,可理解为是配置了大容量存储空间的云计算系统。从三层模型分析,云计算和云存储的访问层、接口层相同,在基础层方面,云存储增加

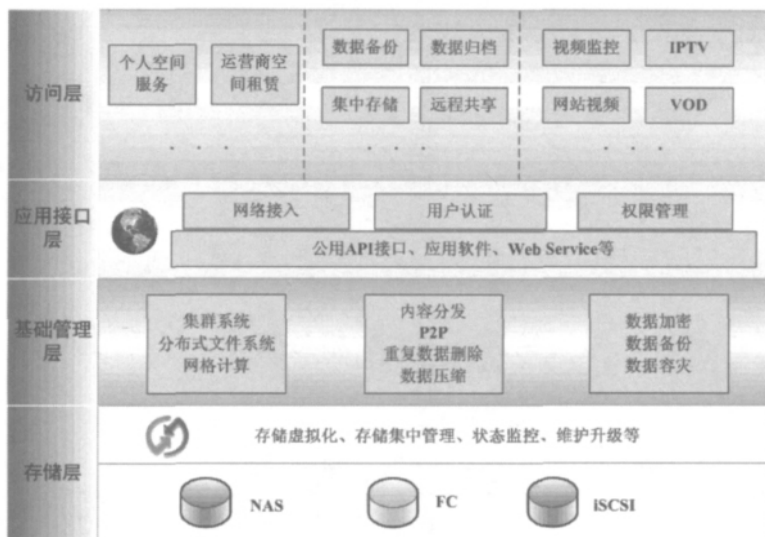


图 1 云存储系统的结构模型

了数据管理和数据安全的部分功能。

3 云存储架构模型

云存储系统的结构模型由 4 层组成,如图 1 所示。

(1) 存储层

存储层是云存储最基础的部分。存储设备可以是 FC 光纤通道存储设备、NAS 和 iSCSI 等 IP 存储设备,也可以是 SCSI 或 SAS 等 DAS 存储设备。数量庞大的云存储设备分布在不同地域,彼此之间通过广域网、互联网或者 FC 光纤通道网络连接。各存储设备上都安装有统一的存储设备管理系统,可以实现存储设备的逻辑虚拟化管理、集中管理、多链路冗余管理以及硬件设备的状态监控和维护升级等。

(2) 基础管理层

基础管理层是云存储最核心的部分,也是云存储中最难以实现的部分。应用接口层通过集群系统、分布式文件系统和网络计算等技术,实现云存储中多个存储设备之间的协同工作,使多个的存储设备可以对外提供同一种服务,并提供更大、更强、更好的数据访问性能。

云存储系统通过集群文件 OS 实现后端存储设备的集群工作,并通过系统的控制单元和管理单元实现整个系统的管理,数据的分发、处理,处理结果的反馈。可利用 CDN 内容分发系统、P2P 数据传输技术和数据压缩技术等保证云存储中的数据可以更有效地存储,使用和占用更少的空间以及更低的传输带宽,从而对外提供更高效的服务。

数据加密技术实现了数据存储和传输过程中的安全性。数据备份和容灾技术可保证云存储中的数据多份保存

不会丢失,保证云存储数据自身的安全和稳定。

(3) 应用接口层

应用接口层是云存储最灵活多变的的部分。不同的云存储运营单位可以根据实际业务类型,开发不同的应用服务接口,提供不同的应用服务。任何一个授权用户通过网络接入、用户认证和权限管理接口的方式来登入云存储系统,都可以享受云存储服务。

(4) 访问层

云存储运营单位不同,提供的访问类型和访问手段也不同。云存储使用者采用的应用软件客户端不同,享受到的服务类型也不同,比如个人空间租赁服务、运营商空间租赁服务、数据远程容灾和远程备份、视频监控应用平台、IPTV 和视频点播应用平台、网络硬盘引用平台,远程数据备份应用平台等。

4 云存储系统拓扑

根据云存储架构的特点,云存储系统应用拓扑如图 2 所示。

云存储系统的核心由云存储控制服务器和后端存储设备两大部分组成。

(1) 云存储控制节点

云存储控制器负责整个系统元数据和实际数据的管理和索引,提供超大容量管理,实现后端存储设备的高性能并发访问和数据冗余等功能。云存储控制服务器是整个系统的统一管理平台,管理员可以在其中监视系统运行情况、管理系统中用户和各项策略等。

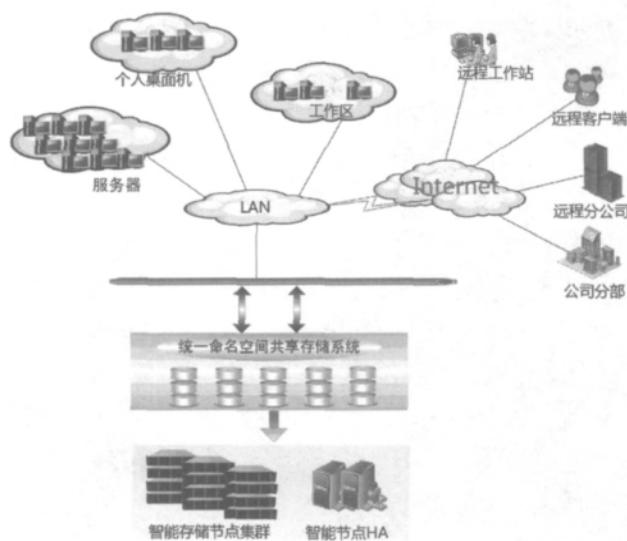


图2 云存储系统应用拓扑

(2) 存储节点

云存储系统采用高性能应用存储设备,可内嵌云存储系统访问协议包、存储节点认证许可等。设备采用高密度磁盘阵列设备,每套设备通过网络接入到云存储系统中,进入云存储存储池后进行分配。对数据存储可实现多副本、多物理设备分别保存,当容量或带宽需要扩展时,通过增加存储节点来实现,根据实际需要灵活扩张,在系统运行中进行在线的容量和性能增加。

5 云存储的技术优势

IDC 研究表明,从 2006 年到 2010 年,全球信息总量增长了 6 倍以上,随着技术的发展,现有传统磁盘阵列中已经普遍采用了 750 GB 或 1 TB 的 SATA 硬盘。但在面对 PB 级的海量存储需求时,传统的 SAN 或 NAS 在容量和性能的扩展上存在瓶颈,已经不能满足新形势下对数据保存高性能、高容量、易扩展的需求。云计算的服务模式也带来云存储的蓬勃发展,对于低成本,但需大容量、高带宽需求下的应用,云存储系统不仅轻松突破了 SAN 的性能瓶颈,而且可以实现性能与容量的线性扩展。相比于传统的集中存储方式,高效集群的云存储系统具有以下几点优势。

(1) 更容易扩容(包括带宽)

当客户需要增加容量时,可按照需求采购(容量或是带宽)服务器,简单增加即可实现容量或者带宽的扩展。如,当需要扩展容量时,采购大容量的存储设备即可;当需要扩展带宽时,采购计算能力较强的服务器设备即可。扩容过程将变得简单:新设备仅需安装操作系统及云存储软件

后,打开电源接上网络,云存储系统便能自动识别,自动把容量加入存储池中完成扩展。相比传统的存储扩容,云存储架构采用的是并行扩容方式,即当容量不够时,采购新的存储服务器即可,扩容环节无任何限制。

(2) 更易于管理

在以往的存储系统管理中,管理人员需要面对不同的存储设备,不同厂商的设备均有不同的管理界面,使得管理人员要了解每个存储的使用状况(容量、负载等)的工作复杂而繁重。而且,传统的存储在硬盘或是存储服务器损坏时,读写效能会降低很多,而且数据很可能丢失,而云存储没有这个困扰,硬盘坏掉,数据会自动迁移到别的硬盘,不需要立即更换硬盘,大大减轻了管理人员的工作负担。

对云存储来说,再多的存储服务器,在管理人员眼中也只是一台存储器,管理人员只要在整体硬盘容量快用完时,采购服务器即可,每台存储服务器的使用状况,通过一个统一管理界面监控,使得维护变得简单和易操作。

(3) 成本更低廉

云存储系统中所采用的存储及服务器设备均是性价比比较高的设备。可实现长久合作关系下,设备商采购的稳定渠道,便于实现对成本及服务质量的控制。

传统的存储系统对硬盘的一致性要求近乎苛刻,必须同厂牌、同容量、同型号,否则系统很容易出问题。面对升级换代较快的 IT 产业,硬盘在使用 2~3 年后很难找到同型号产品更换。使用云存储没有这个问题,云存储系统对存储设备、服务器设备、硬盘等产品的一致性没有要求,不同介质、容量、厂牌、型号的硬盘,都可以一起工作,既可以实现原有硬件的利旧保护投入,又可以实现新技术、新设备的快速更新,合理搭配、优化选择、可持续发展。

(4) 数据更安全,服务不中断

传统存储系统会因为硬件损坏而导致服务停止,如硬盘、主板、电源、网卡等,虽然对此可以设计全冗余的环境(电源、网络、盘阵等),但成本相对太高且工作繁杂。

云存储系统则不同,它可透过将文件和数据保存在不同的存储节点,避免了单一硬件损坏带来的数据不可用。云存储系统知道文件存放的位置,在硬件发生损坏时,云存储系统会自动将读写指令导向存放在另一台存储服务器上的文件,保持服务的继续。

传统存储系统在升级时,往往需要把旧的存储设备文件备份出来后,停机换上新的存储设备,这通常会导致服务的短暂停止。

云存储并不单独依赖一台存储服务器,因此存储服务器硬件的更新、升级并不会影响存储服务的提供,系统会将旧存储服务器上的文件迁移到别的存储服务器,等新的存储服务器上线后,文件会再迁移回来。

6 云存储未来发展趋势

云存储已经成为未来存储发展的一种趋势。随着云存储技术的发展,各家厂商正积极将各类搜索、应用技术和云存储相结合,以便能够向企业提供一系列的数据服务。但从未来云存储的发展趋势来看,云存储系统主要还需从安全性、便携性及数据访问等角度进行改进。

(1) 安全性

从云计算诞生,安全性一直是企业实施云计算首要考虑的问题之一。同样在云存储方面,安全仍是首要考虑的问题,对于想要进行云存储的客户来说,安全性通常是首要的商业考虑和技术考虑。但是许多用户对云存储的安全要求甚至高于它们自己的架构所能提供的安全水平。即便如此,面对如此高的不现实的安全要求,许多大型、可信赖的云存储厂商也在努力满足它们的要求,构建比多数企业数据中心安全得多的数据中心。现在用户可以发现,云存储具有更少的安全漏洞和更高的安全环节,云存储所能提供的安全性水平要比用户自己的数据中心所能提供的安全水平还要高。

(2) 便携性

一些用户在托管存储的时候还要考虑数据的便携性。一般情况下这是有保证的,一些大型服务提供商所提供的解决方案承诺其数据便携性可媲美最好的传统本地存储。

有的云存储结合了强大的便携功能,可以将整个数据集传送到你所选择的任何媒介,甚至是专门的存储设备。

(3) 性能和可用性

过去的一些托管存储和远程存储总是存在着延迟时间过长的问题。同样地,互联网本身的特性就严重威胁服务的可用性。最新一代云存储有突破性的成就,体现在客户端或本地设备高速缓存上,将经常使用的数据保持在本地,从而有效地缓解互联网延迟问题。通过本地高速缓存,即使面临最严重的网络中断,这些设备也可以缓解延迟性问题。这些设备还可以让经常使用的数据像本地存储那样快速反应。通过一个本地 NAS 网关,云存储甚至可以模仿终端 NAS 设备的可用性、性能和可视性,同时将数据予以远程保护。随着云存储技术的不断发展,各厂商仍将继续努力实现容量优化和 WAN(广域网)优化,从而尽量减少数据传输的延迟性。

(4) 数据访问

现有对云存储技术的疑虑还在于,如果执行大规模数据请求或数据恢复操作,那么云存储是否可提供足够的访问性。在未来的技术条件下,此点大可不必担心,现有的厂商可以将大量数据传输到任何类型的媒介,可将数据直接传送给企业,且其速度之快相当于复制、粘贴操作。另外,云存储厂商还可以提供一套组件,在完全本地化的系统上模仿云地址,让本地 NAS 网关设备继续正常运行而无需重新设置。未来,如果大型厂商构建了更多的地区性设施,那么数据传输将更加迅捷。如此一来,即便是客户本地数据发生了灾难性的损失,云存储厂商也可以将数据重新快速传输给客户数据中心。

Cloud Storage Technology

Zhang Longli

(Gammacom Communicate Scheme Design Co.,Ltd., Ha'erbin 150040, China)

Abstract Based on cloud computing and cloud storage concept, analyzes the cloud storage system architecture model, the typical topology, technology advantages and characteristics, and mode of traditional storage architectures were compared, and the future development trend of cloud storage were explored.

Key words cloud computing, cloud storage

(收稿日期:2010-08-06)