

云存储

薛一波¹, 易成岐²

(1. 清华大学 信息技术研究院, 北京 100083;

2. 哈尔滨理工大学 计算机科学与技术学院, 黑龙江 哈尔滨 150080)

[编者按]云存储是随着云计算的产生而发展,并逐渐成为研究热点。云存储是云计算的存储部分,并且可以作为一种服务提供给用户,任何经过授权的合法用户都可以通过网络访问云存储,享受云存储带来的便利。本讲座将分3期对云存储进行讨论:第1期介绍云存储的定义、与传统存储的区别、种类、结构模型及相关标准;第2期介绍云存储的关键技术,包括存储虚拟化技术、分布式存储技术、对等存储技术、存储加密技术、重复数据删除技术、内容分发技术、数据备份技术;第3期探讨目前云存储的成功案例以及云存储的发展趋势。

中图分类号:TN91 文献标志码:A 文章编号:1009-6868(2012)01-0057-04

1 云存储定义

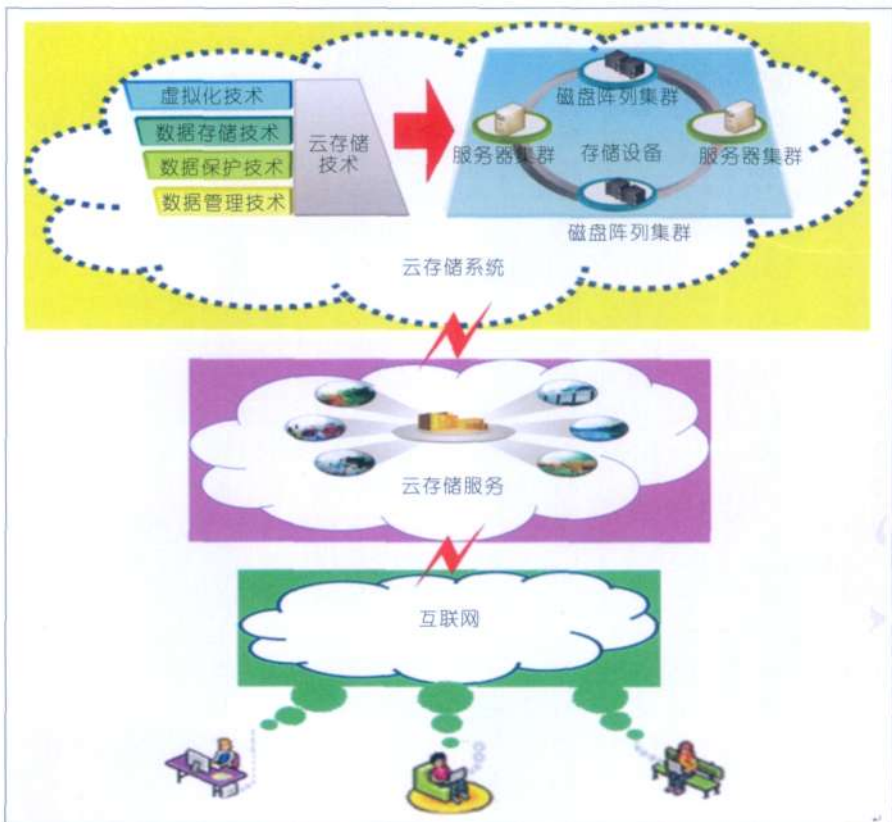
云存储是在云计算概念上衍生、发展出来的一个概念,其目标是使存储空间达到近100%的利用率。云存储除了可以节省整体的硬件成本外,还具备良好的可扩展性、对用户的透明性、按需分配的灵活性和负载的均衡性等特点。近年来,虽然已经有很多公司推出了云存储产品和服务,包括Google的GFS、Microsoft的Azure和Amazon的S3等,但是到目前为止,云存储并没有一个明确的定义。文章引用全球网络存储工业协会(SNIA)对云存储的定义:云存储是通过网络提供可配置的虚拟化的存储及相关数据的服务。

云存储是云计算的存储部分,当云计算处理和运算的核心是大数据量的存储和管理时,云计算就必须部署大量的存储设备。云存储属于云计算的底层支撑,它通过多种云存储技术的融合,将存储设备虚拟化为易扩展、弹性、透明、具有伸缩性的存储资源池,并将存储资源池按需分配给授权用户,授权用户既可以通过网络对存储资源池进行任意的访问和管

理,并按用付费。

图1展示了存储设备、云存储技术、云存储系统、云存储服务之间的

关系。云存储由大量的同构或者异构的存储设备组成,融合了虚拟化技术、数据存储技术、数据保护技术、数



▲图1 云存储系统、云存储技术、存储设备、云存储服务关系图

据管理技术等多种云存储技术,构建一个云存储系统,为用户提供灵活的、方便的、按需分配的云存储服务。

云存储的使用,可以带来如下一些好处:

(1) 从云存储服务提供者的角度看,云存储可以确保方便的管理、低廉的成本和持续运行的服务,这是因为:通过云存储技术可以将不同厂商、型号、位置的存储设备虚拟化成一个存储资源池,云存储服务提供者只需对这一个存储资源池进行管理和分配即可,大大方便了管理,降低了管理并维护了成本;云存储将大量价格低廉的同构或者异构存储设备进行了有效整合,降低了云存储的构建成本;云存储将设备故障和设备升级视为正常情况,采用冗余技术,实现了云存储服务的不间断运行。

(2) 从云存储用户的角度看,云存储可以提供透明的服务和弹性的资源。由于云存储的透明性,云存储用户在使用云存储过程中可以不必考虑存储设备的型号、接口、传输协议等细节问题,大大方便了使用;由于云存储的弹性资源分配,用户可以按需申请存储资源,按使用情况付费,大大降低了使用门槛和使用成本。

近几年,云存储的发展突飞猛进,给整个IT产业界带来了一些很大影响。

首先,云存储改变了产业界的运营模式。过去几乎所有的应用都是安装、运行在用户端,随着云存储的兴起,未来更多的应用可以运行在云端,云存储将传统的软件运营提升到服务运营。

其次,云存储缓解了成本与资源利用率之间的矛盾。传统的数据中心如果想提升资源利用率,必须提高其硬件、网络和管理成本,增加了负担;而云存储可以使用大量的、价格低廉的硬件设备,既降低了成本,又可以通过云存储技术将资源利用率最大化。

再次,云存储促进了绿色节能的

社会责任。随着数据中心规模的日益增长,随之而来的绿色节能问题成为业界关注的焦点。由于传统的存储设备地理位置分散、能源消耗高等问题突出,如何在绿色节能的前提下满足用户的存储需求已经成为一个重要问题。云存储可以将用户的服务进行整合,通过对能耗的统筹管理,实现绿色节能的社会责任。

综上所述,云存储是一种弹性、低成本、高利用率、透明的并能满足用户需求的服务,它采用友好的Web界面与用户进行交互,提供数据存储、数据保护、数据管理等功能,并使用用户身份认证机制来验证用户身份的真实性与唯一性。云存储作为一种新型的应用和运营模式,优势和影响也在逐步扩大,正在引领新的创新和变革。

2 云存储与传统存储的区别

传统存储是指某一具体的存储设备,或者由大量相同的存储设备构成的集合体。当用户使用传统存储时,需要非常清楚地了解存储设备的一些基本信息,比如设备的型号、容量、所支持的协议、传输速度等。此外,对设备的定期维护、软硬件的更新与升级等方面也需要单独考虑。

虽然云存储也是由大量的存储设备组成,但是存储设备可以是异构的,而且对于云存储用户来说,可以不用关心存储设备的基本信息和所在的位置。云存储也充分考虑了设备故障、设备更新和升级等问题,可以提供更可靠的服务。

云存储与传统存储的区别主要如下:

(1) 从功能需求来看

传统存储更关注数据,例如数据的分布式存储、事务处理、数据备份等。由于传统存储的存储方式单一,随着存储设备的更新换代,落后的存储设备将难以处置。存储方式不能随着业务需求的变化而不断变化。改变存储方式时,例如从只读方式转

变为读写方式,必须通过软件的不断更新、甚至重构来解决。

而云存储更关注用户,面向用户提供多种类型的存储服务。云存储具有良好的扩展性,可以使用大量廉价的存储设备,存储方式灵活多样,可以根据业务需求的变化、用户的增减和资金的承受能力,随时调整存储方式。云存储只需对虚拟化后的存储资源池进行统一的管理,即可实现按需使用、按需分配、按需维护。

(2) 从性能需求来看

传统存储对资源的利用率非常低,对存储资源的分配通常是静态的,即参考用户的估计值对存储设备划分成分区或卷,以分区或卷为单位将存储资源分配给用户。由于用户估计值的偏差或者用户需求动态的增减,这样的分配方式会导致一部分存储资源可能长期处于闲置状态,而这些闲置的存储资源无法提供给其他用户。

而云存储对资源的利用率非常高,因为云存储采用动态的方法分配存储资源。另外,云存储对资源的管理也十分的弹性,如果用户的某些资源处于闲置状态,云存储可以将这部分资源进行回收,动态地分配给需要更多资源的用户。

(3) 从成本需求来看

传统存储的投资成本和管理成本都十分昂贵。当使用传统存储时,有时很难提前预测业务的增长量,所以会提前采购设备,很容易造成设备的浪费,存储设备并不能得到完全使用,造成了投资浪费。另外,传统存储的管理员需要管理多种类型的存储设备,不同生产厂商生产的存储设备在管理方式及访问方式又不尽相同,因此管理员需要对各种产品都加以了解,增加了管理的难度及人员的开销。

而云存储可以有效降低投资成本和管理成本。云存储具有很好的可伸缩性、弹性和扩展性,可以灵活扩容、方便升级,由于使用虚拟化技

术,设备管理和维护非常容易。云存储可以根据用户的数量和存储的容量,按需扩容,规避了一次性投资所带来的风险,降低了投资成本;云存储通过存储虚拟化技术,将数量众多的异构存储设备虚拟化,形成统一存储资源池,管理员可以对存储资源池进行统一的管理,最大幅度的降低管理成本。

(4) 从服务需求来看

传统存储容易出现由意外故障而导致服务中止的现象。传统存储将业务和存储相互对应,根据特定的业务划分相应的存储设备。由于存储设备之间的隔离,如果某台设备出现意外故障,业务就会中止,必须将故障修复后才能恢复业务。

而云存储则采用业务迁移、数据备份和冗余等多种技术来保证服务的正常运行,当某个存储设备发生故障时,云存储会根据系统目前的状态,自动将用户的请求转移到未发生故障的存储设备上。发生故障的存储设备恢复后,用户的请求也会重新转移到原存储设备,可以有效地可以保证服务的持续性。

(5) 从便携需求来看

传统存储属于本地存储。数据会保存在本地的存储设备中,并不会和外界进行互联,导致数据具有较差的便携性。

而云存储属于托管存储。云存储可以将数据传送到用户选择的任何媒介,用户可以通过这些媒介访问及管理数据。

云存储在功能需求、性能需求、成本需求、服务需求和便携需求方面都优于传统存储。随着云存储的研究与发展,云存储将逐渐超越传统存储,提供给用户高质量、高标准、高可靠的服务。

3 云存储的种类

目前云存储分为:公共云存储、私有云存储和混合云存储。

(1) 公共云存储

▼表 1 3 种类型云存储比较

类型	公共云存储	私有云存储	混合云存储
可扩展性	高	低	高
安全性	良好(取决于公共云存储采取的安全措施)	最安全(所有的存储都是内部部署)	安全(集成了私有云存储的安全机制)
性能	低等到中等	最高	较高
可靠性	中等	最高	中等到高等

公共云存储是指专为大规模、多用户而设计的云存储。公共云存储的所有组件都建立在共享基础设施上,通过虚拟化技术、数据访问、数据管理等技术对公共存储设备进行逻辑分区,实现对用户的按需分配。公共云存储的存储设备通常设置在用户端的防火墙外部,因此安全性取决于存储设备所采用的安全措施。

因为公共云存储的基本存储设备包括低成本的存储节点及负责跨节点内容分布存储的管理节点,而且公共云存储对用户没有专用存储设备的需求,因此构建成本较低。目前,公共云存储的收费标准平均是每10亿字节收取1美元,但是根据不同的云存储服务提供商或者不同的业务,费用也有所不同。比较典型的公共云存储服务提供商包括:美国电话电报公司、亚马逊、铁山、微软等众多公司。不过,安全性问题及性能问题是公共云存储的核心和焦点问题。

(2) 私有云存储

私有云存储也称为内部云存储,是针对特定用户设计的云存储。与公共云存储不同的是,私有云存储运行在数据中心的专用存储设备上,可以满足安全性及性能的需求。不过私有云存储也存在明显的缺点,就是可扩展性相对较差。因此,私有云存储更适合建立在具有高标准安全性需求与性能需求的数据中心上。目前比较典型的私有云存储系统有:日立数据系统和惠普的 CloudStart 系统。在日立数据系统的私有云存储服务中,用户的数据中心是日立的内容平台,该数据中心由日立公司进行管理,用户需要交纳安装费用,其他费用则按需而付;惠普的 CloudStart 利

用了较为廉价的基础设施,构建出一个全面管理、即用即付的私有云存储产品。

(3) 混合云存储

混合云存储是将公有云存储和私有云存储进行混合的云存储。混合云存储的形式一般情况下以私有云存储为主、公有云存储为辅。虽然私有云存储的出现在一定程度上解决了用户对公有云存储的安全性及性能方面的担忧,但是私有云存储具有需要运行在专用平台上的弊端,使得私有云存储不可能扩展到公有云存储那样的规模。因此混合云存储可以在一定程度上解决公有云存储和私有云存储存在的问题。通常混合云存储需要满足两个条件:一个是整体表现需均衡,即网络延迟需在用户可以忍受的范围之内,并且公有云上的数据应该是无缝透明的;另一个是数据的迁移机制必须完善,即活动的或者经常被访问的数据能够快速从云端提取,不活跃的数据又被推向云端,实现数据的灵活迁移。表1总结了3种类型云存储之间的一些对比情况。

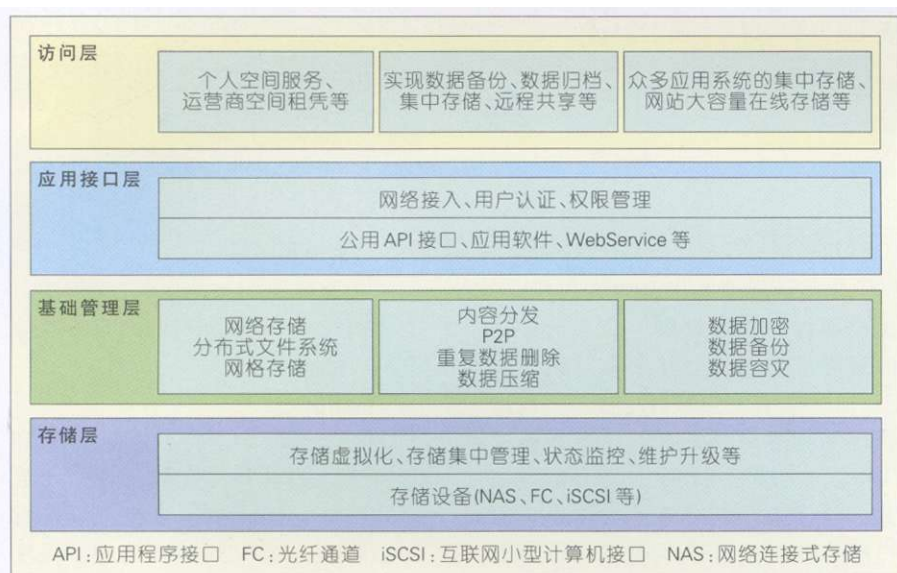
4 云存储的结构模型

在云存储的快速发展中,不同的厂商针对云存储提供了不同的结构模型。目前云存储还没有一个统一的结构模型,文章选取一种比较具有代表性的云存储结构模型,如图2中所示。

这种云存储的结构模型自底向上分为存储层、基础管理层、应用接口层以及访问层。

(1) 存储层

存储层位于云存储结构模型的



▲图2 云存储结构模型

最底层,也是云存储最基础的组成部分,由大量的、多种多样的存储设备构成。在该层中,分布在不同地理位置上的、数量众多的存储设备通过互联网或光纤通道连接起来,构成一个存储资源池。存储层可以实现同构或异构存储设备的虚拟化管理、多链路冗余管理、存储设备状态监控、故障维护、存储设备更新与升级等相关功能。

(2) 基础管理层

基础管理层位于存储层的上一层,是云存储的核心,起到统筹管理的作用。通常情况下,该层也是云存储中最难以实现的部分。基础管理层通过网络存储技术、分布式文件系统、网格存储技术,实现各个存储设备之间的协同工作,从而对外提供同一种服务,实现了强大的数据访问、数据控制、数据管理等功能。该层中的数据加密技术可以保证云存储系统中的数据不会被恶意用户访问或修改,容灾备份技术可以保证云存储系统中的数据不会丢失。

(3) 应用接口层

应用接口层位于基础管理层的上一层,也是云存储结构模型中最灵活多变的部分。该层的主要优点是不同的云存储运营商可以根据实际

需求,开发不同的应用服务接口以应对不同的业务需求,从而提供不同的应用服务。

(4) 访问层

访问层位于云存储结构模型的最上层,是与用户最近的部分。任何经过授权的合法用户都可以通过标准的应用接口登录云存储系统,享受云存储的服务,感受云存储的价值。访问层的构建一般都追求友好化、简便化、实用化的原则。因为云存储服务商的不同,针对用户的业务类型也不同,云存储访问层的访问类型与访问手段也会不同。

5 云存储的相关标准

近几年,云存储的标准化也成为国际上标准化工作的热点之一。为了促进云存储的相关工作能够更好地进行,2009年4月,超过140家公司(包括Cisco、HP、IBM、Sun、VMware、Hitachi、Symantec等)联合成立了SNIA云存储技术工作组。

2009年6月,该工作组发布了一个工作文档《云存储使用场景和参考模型》;2009年9月,SNIA云存储技术工作组根据《云存储使用场景和参考模型》的内容,提出了0.8版本的云数据管理接口(CDMI)规范草案;

2010年4月12日,云存储技术工作组通过对CDMI的修改,提出了CDMI 1.0版本,并且得以通过,最终在网络存储世界峰会(SNW)上公布。

CDMI规范1.0版本的主要有如下一些内容:

(1) CDMI规范可以让不能满足当今需求的、非云存储产品的访问方式演变成云存储访问方式。通过CDMI规范,数据中心对现有网络存储资源的访问可以以一种比较轻松的、透明的方式切换到CDMI云存储资源。

(2) CDMI规范还提供了云存储访问和管理数据的方式。另外,CDMI规范还专注于可扩展性。

(3) CDMI规范将数据对象看成可以用来创建、搜索、更新、删除的独立资源。

(4) CDMI还可以执行管理工作及用于管理型应用程序,可以管理存储的封装、用户的账号、用户的安全访问、监控的信息等。

当然,CDMI规范也有不足之处,就是目前还没有提供一个可靠的评价标准来衡量云存储服务提供商的服务质量是否安全可靠。(待续)

收稿日期:2011-12-15

作者简介



薛一波,中国科学院计算技术研究所、清华大学博士毕业;现任清华大学信息技术研究院研究员、CPU&SOC中心副主任;目前主要研究方向为网络与信息安全、云计算及云安全、计算机体系结构等;承担课题30余项,发表论文100多篇,申请发明专利20余项。



易成岐,哈尔滨理工大学计算机应用技术专业在读博士;主要研究领域为网络安全、云计算、社会计算等;参与过国家自然科学基金、242项目、国家农业部公益项目等多项科研项目。