存储网络虚拟化关键技术的研究与实现*)

刘朝斌 谢长生 张 琨
(华中科技大学计算机学院外存储系统国家重点实验室 武汉430074)

摘 要 存储技术经历了从单个的磁盘、磁带、RAID到存储网络系统的发展历程。存储区域网(SAN)是当今存储网络的主流技术,具有高性能、高可用性和良好的可扩展性等优点。但结构的复杂化不可避免地导致了存储环境与管理复杂度的增加。存储虚拟化是解决存储管理问题的有效手段,本文重点分析了基于存储区域网的存储虚拟化技术,讨论了其关键技术及其主要实现方法。

关键词 存储网络,附网存储,存储区域网,存储虚拟化

The Key Technology and Realization of Storage Virtualization

LIU Zhao-Bin XIE Chang-Sheng ZHANG Kun
(National Storage System Laboratory, Huazhong University of Science and Technology, Wuhan 430074)

Abstract Storage technology goes though the variations from single disk, tape and RAID to storage networking. Now storage area networks (SAN)is the prevalent technology of storage networking, it provides high performance, high availability and high scalability etc., but at the same time, it results in the complications of storage environment and storage management. Storage virtualization is the effective means to resolve the problem of storage management. This paper mainly analyzes the storage virtualization in SAN, discusses its key technology and main implementing methods.

Keywords Storage networking, Network attached storage (NAS), Storage area networks (SAN), Storage virtualization

1 引言

互联网的飞速发展和壮大,导致了许多以数据密集为主要特点的应用:流式多媒体、数字电视、IDC、ASP、ERP、科学计算、数字影像、事务处理、电子商务、数据仓库与挖掘等导致了对存储容量的极大需求膨胀,带来的直接结果就是传统的本地直接存储方式已经远远不能满足这种发展需求[1,2]。当今的网络存储系统正在以超越服务器数倍的速度发展,网络存储市场迅速膨胀,网络存储被推到了IP网络的中央。数据以存储为中心是人们思维的一次重大转变,制造商和提供商以数据为驱动,客户以数据为中心布置业务。

存储技术经历了从单个的磁盘、磁带、RAID,到存储网络系统的发展历程。传统的本地直接存储(DAS)的方式是存储设备附属于某个服务器,数据被局限在某个主机的控制之下,这种方式已远远不能满足企业高可用性、可扩展性、集中统一数据的需要,因而发展出网络存储技术[3,5]。典型的网络存储技术有附网存储NAS(Network Attached Storage)和存储区域网SAN(Storage Area Networks)两种。NAS和SAN各有自己的不同的体系结构、互联协议、文件系统、管理方式等,并在各自的领域得到了很大的应用和发展[4]。然而对于同时购买了这两种类型设备的企业来说,则很难把它们集中到一个单一的系统中去[6]。本文在分析了SAN和NAS各自的缺陷和不足的基础上,详细阐述了存储虚拟化技术的原理和实现层次,分析了基于存储网络的存储虚拟化的关键技术及其主要实现机制。

2 存储虚拟化的实现层次

存储网络为用户带来良好的性能、可扩展性、高可用性的同时,也带来了许多亟待解决的新课题。由于大型企业或公司的关键数据往往分布在几十甚至几百个站点,一个SAN环境中,往往有很多完全异构的服务器和存储设备,因此业务数据的高效、安全管理就越来越受到存储网络界的关注。根据IDC最新数据统计,目前存储管理的成本已经是硬件投资的几倍,而在20世纪80年代,存储管理的成本还只是硬件投资的几分之一。

存储虚拟化技术是解决存储管理问题的有效手段,已经越来越受到业界的关注。通过虚拟化技术,用户可以利用已有的硬件资源,把SAN内部的各种异构的存储资源统一成对用户说是单一视图的存储资源(storage pool),而且采用striping, LUN masking, zoning等技术,用户可以根据自己的需求对这个大的存储池进行方便的分割、分配,保护了用户的已有投资,减少了总体拥有成本(TCO)。更进一步的应用,可以实现SAN与SAN之间的虚拟化、全球的虚拟化。这样从用户角度来看,屏蔽了具体的存储资源的物理细节,不必关心存储设备的配置、物理位置、物理参数甚至容量限制等,这样用户可以专注于自己的业务,增加了用户的投资回报(ROI)。

存储虚拟化并不是一个全新的概念,但是引入到存储领域发生了某些变化,被赋予了新的内涵。存储虚拟化技术包括如下几个实现层次:

2.1 存储设备级虚拟化

最典型的虚拟存储设备是磁盘阵列RAID。RAID的虚拟化是由RAID控制器或控制卡实现的,它提供硬件RAID技术,将多个物理磁盘按不同的分块级别组织在一起,通过板上CPU及阵列管理固件来控制及管理硬盘,解释用户的I/O指令并将它们发给物理磁盘执行,从而屏蔽了具体的物理磁盘,

*)基金项目:本项目得到国家自然科学基金(No. 60173043)和国家863项目(No. 2001AH111011)资助。刘朝斌 博士生,主要研究方向是新型存储系统与结构、高性能存储网络和计算机网络等。谢长生 教授,博士生导师,从事新型计算机外存储体系结构和网络海量存储技术等研究工作。

为用户提供了一个统一的具有容错能力的逻辑虚拟磁盘，用户对 RAID 的操作就像对普通磁盘一样。

2.2 SAN 内部的虚拟化

SAN 内部的虚拟化是现在存储提供商关注的焦点。通过这种虚拟化技术，用户和应用程序可以把 SAN 作为一个单一的、同构的资源池来存取和控制。用户可以根据不同应用系统的需要对这个存储池进行任意地分割并任意地分配给特定的主机或应用系统，从而对 SAN 中的数据进行管理、保护、使用和操作，对设备进行监控，充分利用了 SAN 的存储能力。SAN 内部的虚拟化的实现，需要在原有体系结构中加入一个新的虚拟化层架构。通过这个虚拟化层，为应用程序和用户提供 SAN 的全局虚拟化视图，从而从用户和应用程序的角度看来，原来复杂结构的 SAN 就是一个结构相对简单的、具有统一界面的虚拟存储池，它对用户和应用程序完全透明，而存储池中逻辑存储单元的具体细节则只是系统管理员所关心的问题。

2.3 广域(全球)虚拟化

由于经济全球化和信息网络化的直接影响，越来越多的企业或公司的业务不断向全球拓展，企业的关键数据往往分布在许多个 SAN 中，如何高效地管理这些数据已经是困扰业界的重要问题。随着 IP 存储的兴起，FC SAN 和 IP 存储结构之间的联系越来越紧密，IP 存储可以作为 FC SAN 的重要补充。多个独立的 SAN 可以通过 IP 网络连接成一个统一的存储网，而存储虚拟化也是实现全球统一存储网络的一种重要技术手段。

广域的存储虚拟化是存储资源抽象的最高的形式，它将全球范围的存储资源统一成一个巨大的逻辑存储池。它的出现将使大规模存储和计算彻底分离，导致了存储服务提供商(SSP)和存储服务代理(SSA)的诞生。实现了广域的存储虚拟化，用户的计算机最终将只需要计算功能和传输功能，当需要使用大规模数据存储时，可以向存储服务代理提出请求，后者将从广域的虚拟存储池中选择合适的存储资源，并向相应的存储服务提供商提出具体的存储需求。存储服务提供商将按需分配相应存储资源给用户，并提供各种功能服务。全球存储资源虚拟化将是未来存储工业界的终极目标。

3 存储虚拟化网络的关键技术

存储虚拟化的核心工作是物理存储设备到单一逻辑资源池的映射，通过虚拟化技术，为用户和应用程序提供了虚拟磁盘或虚拟卷，而且用户可以根据需求对它进行任意分割，并分配给特定的主机或应用程序，而为用户隐藏或屏蔽了具体物理设备的各种物理特性。

在一个存储网络中，经常会有不同类型的服务器，不同的存储设备，不同的接入方法及不同的接入协议(如：FC，ESCON，iSCSI，SSA，Infiniband)等，由于存储网络的复杂性，就增加了实现的难度，下面着重探讨一下实现存储虚拟化的一些关键技术。

共享冲突与数据一致性 存储虚拟化的一个主要功能是实现存储数据的共享，普通的文件系统只允许对数据进行独占式访问，但是商业应用需要在操作系统和“数据仓库”之间共享数据。数据的不同拷贝应能在不同平台上的应用程序之间传输。此时必须防止不同服务器操作系统所带来的存储共享冲突和并行存储时的 I/O 访问冲突等。这就需要良好的锁机制算法、多种级别的锁机制以及 Cache 一致性等技术，来保证数据之间的连贯性和一致性。

异构性 实现真正意义上的设备互操作性，简化在由不同主机操作系统和不同设备类型组成的异构存储环境中的系统管理和用户操作。随着全球经济化与信息化的发展，不仅仅是广域范围的设备环境复杂，即使是在同一个 SAN 内都可能出现使用不同操作系统的服务器和来自不同厂商的存储设备，这种复杂的环境就要求虚拟化软件必须对异构环境具有极强的适应性。要解决这个问题，首先应该尽快制定一个操作平台厂商、网络厂商和存储设备厂商都必须遵循的标准，对于存储虚拟化，则要选择一个好的虚拟化系统结构，具有一定的灵活性和自适应性。

数据安全 将传统的附属于某个主机的存储模式转为现在的存储网络，在带来高性能、高可用性等优点的同时，我们也应该清醒地认识到，大量数据在广域网上传输，安全性肯定有所降低，存储网络的出现使得外部网络有机会访问网内的存储设备，从而增加了数据被越权访问和恶意攻击的危险性，因此如何保证数据的安全是虚拟化技术要面临的课题之一。同时还要研究系统的故障检测、故障隔离和故障恢复技术。

全局管理 大多数存储管理系统是以服务器为中心的或以任务为中心的，这种局部范围内的分散的存储管理接口导致了对资源的低效使用。企业需要在不中断正常营业的情况下对存储容量和规模进行无限扩充，此时新的存储设备应被透明地加入系统中，即从一个中心平台自动发现、监测和管理任何厂商的 SAN 设备。同时存储池需要高效的管理以免浪费资源，存储容量应被均匀地分配给各个服务器上的应用程序以优化负载等都是虚拟化需要解决的关键技术。

基于策略和用户按需存储 不同的需求对存储的要求是不同的，而不同的存储设备也有不同的性能，因而要研究如何将不同的存储资源分配给最适合它的应用。可以按照不同的应用，不同的性能，不同的服务优先级等各种策略来为用户提供定制服务，满足不同用户和应用程序的需求，实现真正意义上的按需存储和个性化存储。

4 存储虚拟化的实现模型及方法

4.1 虚拟化技术的实现模型

根据 SAN 中数据与控制信息是否使用同一通道，SAN 可以分为对称结构和非对称结构两种实现模型。对称结构又称为 in-band 模式，如图1所示。这是普通 FC SAN 最常用的形式，数据和控制信息使用同一条通路，节省了硬件投资，但是容易造成网络拥塞，降低了性能，同时容易产生瓶颈和单点失效，故在应用中这种结构往往是冗余配置。

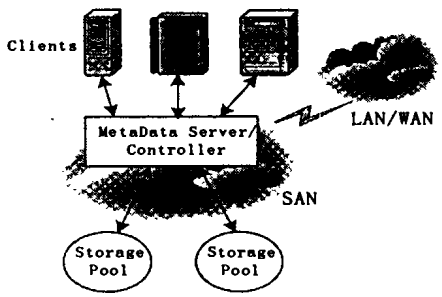

图1 对称结构

非对称结构又称为 out of band 方式，数据和命令信息使用不同的通道(如图2所示)，应用服务器的 I/O 命令先通过专用的命令通路传送至专用的元数据服务器或控制器，获得元数据和数据视图后，再直接通过数据通路得到所需要的数据。由于数据在专用的通道上传输，因此提高了性能，且避免了单点故障和瓶颈，但是在一定程度上增加了用户投资。

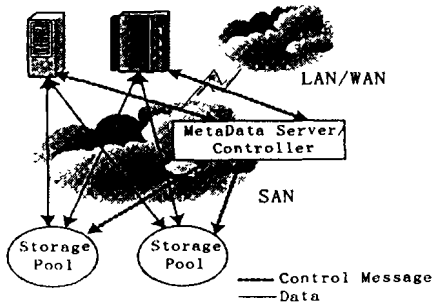

图2 非对称结构

4.2 存储虚拟化实现方法

SAN的虚拟化是通过虚拟抽象层来实现的。它将系统的实际存储设备的物理布局和结构等存储资源整合起来，分成虚拟磁盘或虚拟卷，采用虚拟卷，使得应用程序和用户就像使用本地磁盘一样，呈现给服务器操作系统的是一种物理磁盘的抽象。用户不但可以根据预先设定的策略和需求对其进行任意分配，而且从理论上允许任意地以单个磁盘为单位进行在线扩展。基于存储区域网的虚拟化将会取代传统的服务器受限的存储管理体系结构，并且将进一步影响存储器硬件的结构与发展。虚拟化SAN有多种实现方式，根据虚拟化层在存储网络体系结构中的实现位置，可以划分为基于主机、基于存储设备和基于存储网络的虚拟化。

(1)基于主机的方案是将虚拟化层放在SAN的应用服务器上，通过改造操作系统或者加上一层虚拟层来实现映射工作。这种方法不需要额外的特殊硬件，虚拟化层以软件模块的形式嵌入到应用服务器的操作系统中，为连接到SAN上的各种存储设备如磁盘、阵列等提供必需的控制功能，这样，就使得主机的操作系统在运行应用程序的时候就好像与一个单一的存储设备直接通信一样。这种方法有其自身不可避免的缺点：首先，兼容性不好，由于虚拟化层驻留在应用服务器上，因而软件模块就必须能嵌入到各种类型的操作系统中，因此，这种方法往往适合于配置在同一个厂商的服务器，甚至是一个同构的存储环境中，这就增加了用户的设备依赖型和局限性。其次，这种虚拟化的技术实际上是在一个分布式的环境中实现的，当SAN中的任何一个欺诈主机、人为错误以及操作系统异常等出现时，就可能会影响到所有连接到SAN的存储设备的数据的完整性和一致性，因此需要适当的集中统一管理策略。

(2)基于存储设备的实现方法受到RAID技术的启发，是将虚拟化层放在存储设备上的适配器、控制器等来实现的。这种方法简单，易于实现，且将SAN中应用服务器的存储化工作分离出来，直接在存储设备实现，极大地提高了性能。但是由于目前SAN的虚拟化技术还没有一个统一的国际标准，各个提供商为了巩固自己的市场份额，各个厂家各自为政，都是根据自己的存储产品特点来开发虚拟化，因而不同厂家的存储产品很难在同一个SAN中无缝连接，达不到预期的性能，且调试成本昂贵。

(3)基于存储网络的虚拟化是目前SAN虚拟化的主流技术，通过在存储区域网络这一级采用智能化的路由器、交换机，或者是增加一个元数据服务器等来实现虚拟化的工作。

基于交换机的虚拟化方式为许多厂商所采用，通过改造或添加中间件的微代码，使得交换机同时完成交换功能和虚拟化功能。由于网络中的交换功能和虚拟化功能同时在同一个设备上完成，结合紧密，因而大大改善了系统性能。另外，为了提高交换机的性能，还普遍地采用了缓存技术和优化的缓存调度算法。

基于存储路由器的虚拟化方式是一种比较新的趋势，它是随着SAN存储路由技术的出现而发展起来的。与存储网络中的交换机相似，存储路由器是一种智能化的设备，它既具有普通的路由功能，又针对存储网络的特点增加了虚拟化的功能，并且能够完成协议转换，从而连接不同的存储网络。通过改造路由器或添加中间件，使得虚拟化功能和自身的路由功能紧密结合起来，数据传输和控制信息使用同一个通路，拓宽了应用领域。采用这种方法，可以结合系统的实际情况而灵活地采用对称结构或非对称结构，虚拟化层不需要在应用服务器上增加软件模块，减少了主机的负载和复杂度；同时采用这种路由器接入，不但可以为SAN中的存储设备提供虚拟化功能，而且还可以将通用的以太网上的大量普通用户按照需要的策略连接到虚拟存储池中，在存储网络环境中，采用多个路由器，分布式地存放元数据和全局逻辑视图，还解决了单点失效和瓶颈问题。但是，采用这种方法，用户不得不购买路由器以接入存储网络，不可避免地增加了用户投资。

基于专用服务器的虚拟化方式是目前一种被普遍采用的配置架构，多年来，由于人们习惯于专门配备文件服务器、数据库服务器等，自然就想到了在SAN中采用一台特殊的服务器，一般称为元数据服务器(metadata server)或元数据控制器，专门用于提供存储虚拟化功能。在这种解决方案中，一般在应用服务器上，还驻留着一个很小的虚拟化的代理软件模块，用于维护本地的数据视图和重定向I/O。元数据服务器负责管理整个存储网络环境中的数据管理工作，提供了虚拟化的平台，并驻留着整个虚拟存储池的全局数据视图。当应用服务器所需的数据在本地数据视图上时，就不经过元数据服务器而直接存取；若应用服务器所需的数据不在本地的数据视图上时，主机上的虚拟代理将把I/O发往元数据服务器，元数据服务器就把它维护的相关的数据视图和元数据返回给主机，主机再去存取数据并将相关的数据视图写到自己的本地视图中，这样，当下次访问时就可以不经过元数据服务器而直接存取，从而大大提高了性能。

结束语 本文从分析传统存储技术入手，分析了传统存储演变到存储网络的必然。重点分析了存储网络中基于SAN的存储虚拟化技术，探讨了存储虚拟化的关键技术和主要实现方法。

正如个人PC和Internet的迅速发展得益于广泛的开放性一样，通过采用存储虚拟化技术，SAN正在通过开放系统的存储资源让信息变得更加开放。SAN打破了存储器与服务器之间的束缚，允许你独立地选择最佳的存储资源或者是最佳的服务器，从而提高了可扩展性和灵活性。存储虚拟化的研究与发展必将极大地推动存储网络的前进步伐。

参考文献

1 Arunkundram R S, Sachdev P, Karapagam P. Special Edition Using Storage Area Networks. Que, 2001.11
2 Liu Zhaobin, Xie Changsheng, Fu Xianglin, Cao Qiang. A high scalable and performance storage architecture for multimedia applications. In: Proc. of SPIE, v 4861, 2002.116～120
3 Khattar R, Murphy M, Tarella G, Nystrom K. Introduction to Storage Area Network. Redbooks Publications (IBM), SG24-5470-00, Sep. 1999
4 Phillips B. Have Storage Area Networks Come of Age? IEEE Computers, 1998.7
5 Clark C T. The Virtualization of Storage. TidalWare white paper
6 Farley M(美)，孙功星等译. SAN存储区域网络. 机械工业出版社，2002.4