

# 虚拟网络存储空间划分与抽象技术研究<sup>\*</sup>)

刘朝斌<sup>1</sup> 吴非<sup>2</sup> 胡迪青<sup>2</sup>

(大连海事大学计算机科学与技术学院 大连 116026)<sup>1</sup>

(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)<sup>2</sup>

**摘要** 虚拟网络存储正在成为存储领域研究的热点,然而现有的虚拟化技术大多缺乏具体有效的空间划分与抽象技术。本文在分析网络存储特点的基础上,提出了网络存储的总体抽象拓扑结构模型,给出了一种虚拟网络存储空间的划分与映射方式,并研究了客户端和服务端端的虚拟 I/O 重定向技术,最后探讨了虚拟存储服务器的性能模型。

**关键词** 存储虚拟化, 网络存储, 存储空间, I/O 重定向

## Research on Space Partition and Abstract Technology for Storage Virtualization Networking

LIU Zhao-Bin<sup>1</sup> WU Fei<sup>2</sup> HU Di-Qing<sup>2</sup>

(Dept. of Computer Science and Technology, Dalian Maritime University, Dalian 116026)<sup>1</sup>

(National Storage System Laboratory, Huazhong University of Science and Technology, Wuhan 430074)<sup>2</sup>

**Abstract** Virtual network storage is becoming the hotspot in the storage research area. However, the traditional storage virtualization technologies usually lack effective space partition and abstract realization. Based on the analysis of the characters of network storage, this paper introduced a general abstract topology model for network storage and presented a space partition and mapping method for storage virtualization. We also researched the I/O redirection of clients and sever, and then discussed the performance models of virtual storage sever.

**Keywords** Storage virtualization, Network storage, Storage space, I/O redirection

## 1 引言

随着 Internet 与存储等相关技术的迅猛发展,各大 Internet 服务提供商(ISP)以及企业公司的文件服务器、数据库服务器的数据服务能力越来越力不从心,并且这种发展速度令人根本无法预测未来几年后的数据量会增长多少<sup>[1]</sup>。因此,如何能够保证一个存储系统具有高性能的 I/O 吞吐率、高可靠性和高可扩展能力,以及良好的容错性能,成为各个 IT 厂商倾注极大热情去解决的重大问题<sup>[2]</sup>。就目前看来,解决的方案不约而同地采用了能够提供高性能价格比的存储网络技术。

附网存储 NAS(Network Attached Storage)技术是网络技术在存储领域的延伸和发展<sup>[3]</sup>。它直接将存储设备挂在网上,具有良好的共享性、开放性、可扩展性,但缺点是与 LAN 共用同一物理网络,易形成拥塞而影响 I/O 性能,特别是在做数据备份等应用时性能较低,影响了它在企业级存储中的地位<sup>[4]</sup>。

存储区域网 SAN(Storage Area Networks)技术的存储设备是用专用网络相连的,这个网络是一个基于光纤通道协议的网络。光纤通道的存储网和 LAN 分开,且具有高可靠性、高传输率的特点,因此系统性能就得到有效的提高<sup>[5]</sup>。由于具有这些优异的性能,所以 SAN 成为企业存储的重要技术。然而,由于传统的 SAN 技术一般是采用光纤通道技术互联,而基于 FC 的互联设备如集线器、交换机等的价格居高不下,高出以太网互联设备好几倍,从而造成用户投资比较高,而且很多用户没有成熟的技术,也造成了培训和管理的成本额外增加<sup>[6]</sup>。另一方面,就近几年的发展和应用来看,SAN 系统

的互操作性问题一直是困扰其发展的重要障碍,不同厂家的存储设备很难互联到一个统一的 SAN 中。

针对以上对 SAN 和 NAS 缺点和不足的分析,结合相对成熟的 IP 技术的迅速发展和基于 IP 技术的产品和设备的日益丰富,因而就提出了虚拟网络存储技术<sup>[7]</sup>。即采用成熟的 TCP/IP 技术构成存储网络,提出了一种兼备 NAS 和 SAN 优点的统一存储网。这样就可以充分地利用用户现有设备资源和成熟的 IP 管理技术。

虽然虚拟网络存储受到了业界人士的广泛关注,但是目前大多还停留在探索阶段。本文给出了一种虚拟网络存储的存储空间划分方法,研究了客户端和服务端端的虚拟抽象技术,并对系统核心组件——虚拟存储服务器进行了重点研究。

## 2 网络存储拓扑抽象结构

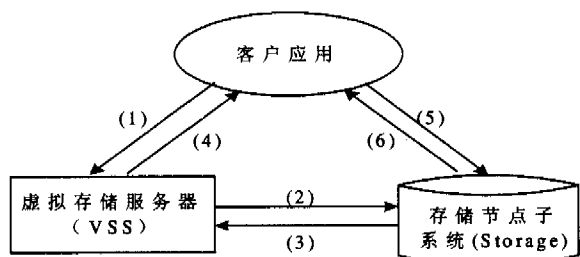
一般的虚拟网络存储系统都包含传统的直接存储(DAS)、磁盘阵列(RAID)、通常的网络存储系统结构(如 SAN、NAS、iSCSI 等)<sup>[8]</sup>。虽然它们的传输协议、存储容量、内部组成、I/O 流程等可能各有不同,但是组成一个大型的网络存储系统后的基本组成元素是一致的,因此我们都可以把它们的实现归纳为一种统一的抽象拓扑结构,如图 1 所示。

该抽象系统包含三个主要节点组件:上层客户端应用、虚拟存储服务器(VSS)和虚拟存储节点子系统。当然,在实际配置中,一台物理设备或一个系统节点都可能同时扮演着其中的两种甚至三种角色。从图 1 中可以看出系统的 I/O 服务流程:在实际应用中,上层客户端向系统提出读写请求,首先到达 VSS,VSS 通过与存储子系统的交互工作,以及对系统元数据的维护查询,返回给客户端元数据的详细信息,并建

<sup>\*</sup> 基金项目:本项目得到国家自然科学基金(No. 60173043)资助。刘朝斌 博士、副教授,主要研究方向是新型存储系统与结构、高性能存储网络、网络计算和分布式文件系统等;吴非 博士生、讲师,主要研究方向为计算机系统结构;胡迪青 博士、副教授,主要研究方向为计算机系统结构。

立客户与存储设备的 I/O 对话连接。此时,上层应用就可以直接和相应的存储设备进行 I/O 访问,实施传输行为。同时,由于客户请求可以包含存储 I/O 策略选择,因此客户应用可以选取优化 I/O 性能的存储模式进行读写等 I/O 传输操作。

相应地,对应于虚拟存储的抽象拓扑结构,我们也分别有针对性地对这三个组件的虚拟化支撑技术,即客户端存储虚拟化、VSS 存储虚拟化和存储子系统的 I/O 虚拟化。



- (1)客户端向系统提出请求;
- (2)VSS 查询存储节点;
- (3)存储节点向 VSS 返回状态等信息;
- (4)VSS 将控制信息返回给客户端;
- (5)客户端根据控制信息,与相应的存储节点建立连接;
- (6)存储节点与客户间进行存取过程。

图1 虚拟网络存储的统一抽象拓扑结构

### 3 虚拟网络存储空间划分与抽象

#### 3.1 存储空间划分

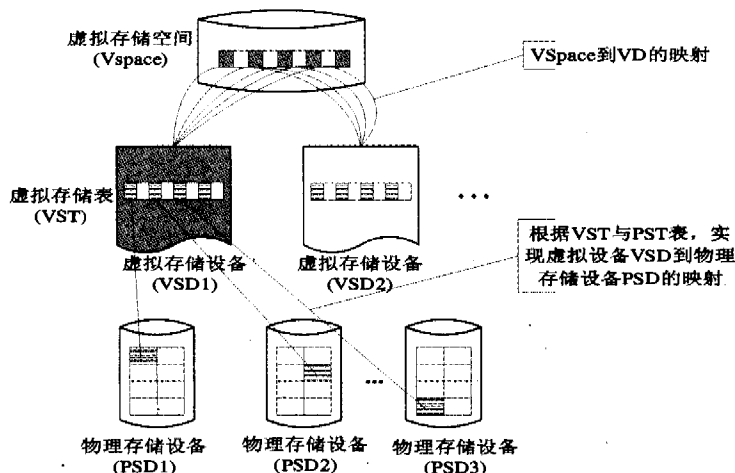


图2 虚拟存储空间的一种划分与映射

在虚拟网络存储中,存储资源的属性分为两级视图:一种是用来表示内部数据的物理存储资源表;另一种是呈现给全局用户的虚拟存储表。图2所示为PST与VST的一种映射,在分布式虚拟存储环境中,这种映射组织与管理由虚拟存储服务器实施。通过这种映射,用户的I/O请求通过系统策略分配,可以自动地定位到具体的物理存储设备上,而对上层应用透明。

#### 3.2 虚拟存储抽象技术

为了实现存储虚拟化技术,在客户端和服务端还必须采用相应的虚拟化技术。为此,我们研究和开发了基于文件系统层的I/O重定向技术。

客户端虚拟代理不必保存任何存储数据的元数据,运行在客户系统,提供对远程数据及其元数据的访问,如图3所示。将远程的文件系统资源与本地的文件系统资源视图统一起来,为上层应用提供服务(虽然在细节上对远程资源的访问与访问本地资源可能并不完全一样),而且增加了虚拟代理模

在虚拟网络存储中,有很多不同特性的存储设备节点,它们的内部组成、存储容量、传输介质等都可能各有不同。为了把它们有机地虚拟整合在一起,我们把不同节点上的物理存储设备(PSD)用物理存储资源表(PST)来描述。相应地,把一组逻辑上相关的存储设备定义为虚拟存储设备(VSD),并用虚拟存储表(VST)来描述和管理,而系统所有的虚拟设备就形成了整个系统对外的虚拟存储空间(Vspace),如图2所示。一个虚拟存储设备可能包含一个或几个物理存储设备,或者是物理设备的一个或几个分区,或不同的逻辑卷(LUN),甚至细化到逻辑分区上的目录。例如图2中,虚拟存储设备VSD1分别由底层具体的逻辑卷构成,而这些逻辑卷位于不同的物理存储设备上(PSD1-3),而不同的VSD则共同组成系统上次的虚拟存储空间,对上层用户提供透明的服务。

关于VSD的划分有很多种,可以根据小写I/O、流媒体文件、目录文件等来划分;也可以根据存储设备的性能、速度、带宽等来划分。同时,VST和PST的定义也非常关键,涉及到实现技术的效率和效能,也会最终影响到整个存储系统的I/O性能。因此,我们定义了非常详细的VST和PST属性表,包含了各种可能情况下的数据结构和元数据结构。比如对虚拟存储节点,定义了全局ID、节点名、节点IP以及节点类型等;为了定义I/O存储策略,定义了线性叠加、镜像、分条和自适应等机制。篇幅所限,此处不再一一列出。

块后,并不会影响到客户端的本地I/O操作。

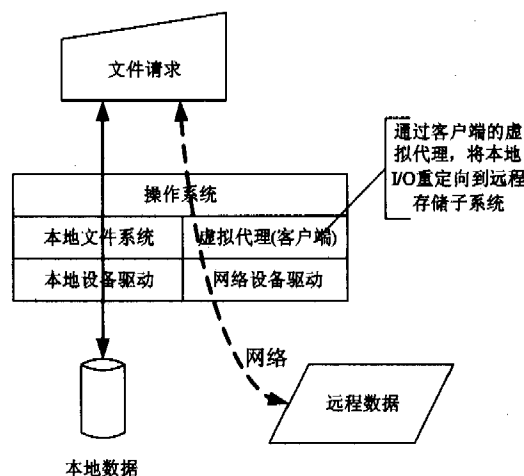


图3 客户端I/O虚拟抽象图

虚拟存储服务器的虚拟抽象技术如图 4 所示。对于原有的本地 I/O 请求,系统并未做任何改动,然而对于远程的网络 I/O 进行了重新设计。服务器从 NIC 接收远程的 I/O 请求,通过服务器端的虚拟化代理传送到本地文件系统和卷管理器,进而映射到数据块级的设备,并将定位后取得的数据信息通过路由返回给远程客户。在此过程中,服务器的主要操作包括登陆和进程路由、身份认证、进程调度、I/O 策略调度以及安全控制等。

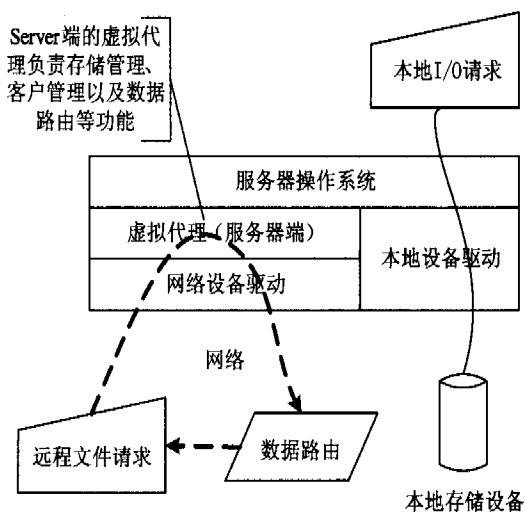


图 4 虚拟存储服务器的 I/O 虚拟抽象图

#### 4 虚拟存储服务器性能模型

显而易见,虚拟存储服务器(VSS)是整个虚拟存储网络系统的核心部件,因此研究存储服务器的实现技术和性能模型具有重要的意义。

如果用  $T$  代表整个系统的性能,那么  $T$  可以用一个四元组来表示:

$$T = T(T_{VSS}, T_{Vnode}, T_N, T_O)$$

其中,  $T_{VSS}$  是虚拟存储服务器的性能参数;  $T_{Vnode}$  是虚拟存储节点的性能参数;  $T_N$  是网络性能参数;  $T_O$  是其他性能参数。

显然,虚拟存储服务器对整个系统影响巨大。它负责管理整个系统的资源、分配 I/O 请求、维护元数据信息和确定 I/O 策略的选择等。另外,服务器对数据备份与恢复、数据一致性机制、系统配置以及冗余校验起着决定性作用。更高一级的功能是可追踪记录文件系统的变化,为保持映像、备份及同步的一致而动态地对其它系统组件进行调整。VSS 的主要设计目标与功能概括如下:

##### 1) 元数据管理

VSS 的主要功能是对系统元数据的管理与维护。元数据是描述数据属性的数据,如访问模式、内部组织、锁机制等,包括应用需要知道数据存放在哪个节点的存储设备上,以及在这个设备上的起始位置的相关信息。同时,VSS 还负责虚拟存储设备表(VST)和物理存储设备表(PST)的更新与维护,以及物理块地址与逻辑块地址相互转换的内部数据结构信息。另外,由于设备在从一种系统移植到虚拟网络存储系统的过程中名称可能会发生部分变化或很大的不同,因此 VSS 必须采取额外的措施来重新集成现有的文件系统和卷。

##### 2) 存储设备管理

虚拟存储网络环境中各种各样的存储介质,VSS 要负责管理整个系统下的全局设备,包括硬盘、磁带驱动器和光驱等。为设备分配全局名称,并负责维护到虚拟存储设备表

(VST)中去,设备也可在系统启动后动态添加到这个设备管理表中。同时我们注意到,不同的存储设备有不同的存储特征,包括容量、速度和 I/O 性能等。磁带库的价格一般比较便宜,但是容量一般很大,不过转速较慢。而且,即使是同一提供商的产品,也有很大不同。因此,对存储设备,我们进行了经济而有效的分组设计,使之更有效地匹配不同特点的存储 I/O,同时兼顾整体 I/O 性能。

##### 3) 虚拟节点管理

存储网络不同于单机的存储模式,在大存储网络环境下,存储子系统是分布式的,有很多存储节点。VSS 的主要功能之一就是研究这种分布式的网络存储管理,系统将用户在网络上分布的、异构的存储服务器有机地组织起来,为用户提供一片可扩展、完整、大容量的虚拟存储空间,同时提供可靠、透明的存储应用服务。如网络用户通过任意一台局域网上的计算机,都可以访问到虚拟存储空间,用户可以实现应用程序的自动迁移和数据的自动备份与恢复等。

##### 4) I/O 策略管理

为了优化存储 I/O 性能,我们采用了许多 I/O 策略。比如对于不同的请求大小,我们可以决定并行的粒度;对于虚拟节点性能的不同,可以分配数据分条的频度;以及访问模式与节点模式匹配策略等。

**结束语** 计算机和网络高度发达的时代,数据对企业和个人尤为重要。伴随着网络存储的不断发展,需求与技术的自然选择使网络存储进入到了高级阶段——存储虚拟化。随着生产力的提高,人们需要管理越来越多的数据,靠传统手工人力管理海量数据已经不可能。而虚拟存储技术的提出是一项战略性的全新存储理念,它彻底变革了存储部署、应用与管理方式。

本文给出了网络存储环境下的虚拟存储的一种总体抽象拓扑结构模型,提出了虚拟网络存储空间的一种划分与映射方式,并分别给出了客户端和服务器的 I/O 重定向虚拟技术,探讨了虚拟存储服务器的模型和主要功能。在后续计划中,我们还要继续研究虚拟网络存储的自适应性、安全性以及并行对存储 I/O 性能的影响。

#### 参 考 文 献

- 1 Morris R J T, Truskowski B J. The evolution of storage systems [J]. IBM Systems Journal, 2003, 42(2): 05~217
- 2 Pollack D. Practical Storage Area Networking [M]. Addison Wesley, 2002. 6~19
- 3 Bassi J S, Moor A. Managing Data Storage in the Network Plank [J]. IEEE Internet Computing, 2000, 15(5): 50~58
- 4 Zhang Hong, Koay Teong Beng, Pallayil V, et al. Fibre Channel Storage Area Network Design for an Acoustic Camera System with 1.6 Gbits/s Bandwidth. Electrical and Electronic Technology [J]. TENCON, 2001(1): 143~148
- 5 Liu Zhaobin, Xie Changsheng, Wu Fei, et al. High-Speed Interconnection for Storage Area Networks. In: Proceedings of SPIE, v5060, 2002. 366~369
- 6 Bhide A, Engineer A, Kanetkar A, et al. File Virtualization with the DirectNFS Protocol. In: The 19th IEEE Symposium on Mass Storage Systems, 2002. 1~6
- 7 Milligan C, Sellkirk S. Online Storage Virtualization: The Key to Managing the Data Explosion. System Sciences. 2002. HICSS. In: Proceedings of the 35th Annual Hawaii International Conference on, 2002. 3052~3060
- 8 刘朝斌. 虚拟网络存储系统关键技术研究及其性能评价: [博士学位论文]. 武汉: 华中科技大学, 2004