

网络存储的统一与虚拟化^{*})

张成峰 谢长生 罗益辉 罗东健

(华中科技大学计算机学院外存储国家重点实验室 武汉 430074)

摘要 存储资源共享是网络存储的最基本的任务,为了给用户提供有效的存储资源共享,网络存储的统一化和虚拟化采取了不同的方法:统一化的侧重点在于存储技术、存储资源的整合和存储管理的研究,虚拟化侧重于如何为用户提供更方便的应用环境。统一虚拟存储体系结构则结合了两者的优点,代表着网络存储的将来。本文对网络存储的统一化和虚拟化的现状和存在问题进行了探讨,分析了阻碍统一虚拟存储体系结构实现的不利因素,针对这些存在的问题,本文给出了一个基于自主式存储中介的网络存储系统,该系统结合了网络存储的统一化和存储虚拟化技术,可以看作统一的虚拟存储网络的一个实现。

关键词 NAS和SAN的融合,存储虚拟化,统一虚拟存储

The Integration and Virtualization of Networking Storage

ZHANG Cheng-Feng XIE Chang-Sheng LUO Yi-Hui LUO Dong-Jian

(National Storage System Laboratory, School of Computer Science, Huazhong University of Science and Technology, Wuhan 430074)

Abstract The share of storage resources is the basic duty to the network storage. To effectively supply the share of storage resources for the clients, the technologies of network storage integration and that of network storage virtualization are used. The integration of networking storage laid heavy stress on the storage technologies, integrating methods and storage management, while the storage virtualization laid heavy stress on the improvement of the application environment for the clients. Having the advantages of the integration and the virtualization, the universal virtual storage architecture will be the future of networking storage. The technologies of both the integration and the virtualization are introduced in this paper. To realize the universal virtual storage architecture, we still have a lot of job to do. A network storage system based on ASAS(Autonomic Storage Agency System) is introduced, which combines the technologies of the integration and the virtualization, and it can be seen as a kind of methods to realize the universal virtual storage architecture.

Keywords Integration of NAS and SAN, Storage virtualization, Universal virtual storage

据有关机构预测,至少在5年内,存储市场都会以10%~20%的年增长率递增,其中网络存储部更将以每年增长62%的速度高速发展。对网络存储产品的旺盛需求,极大地推进了网络存储技术的发展,网络存储的统一化和虚拟化是其中的两个热点。

目前,存储网络的统一主要是指NAS和SAN的融合。NAS和SAN是网络存储的两种不同的体系结构。在实际应用中,它们各有特点,也有各自的局限性。NAS和SAN的融合,可以较好地解决NAS和SAN之间的存储共享问题,有效地利用各自的特点,提高可扩展性、灵活性以及效率。

存储虚拟化是指将具体的存储设备或存储系统同服务器操作系统分隔开来,为存储用户提供统一的虚拟存储池。存储虚拟化是具体存储设备或存储系统的抽象,展示给用户一个逻辑视图,同时将应用程序和用户所需要的数据存储操作和具体的存储控制分离。存储虚拟化能够极大地方便用户对存储资源的使用,减小存储系统管理开销,优化存储系统性能,提高存储资源利用效率。当前,基于网络的存储虚拟化的主流是SAN的虚拟化。

随着网络和计算机技术的发展,新的网络存储技术和标准,不断发展地出现,怎样有效地整合不同存储网络将是极具挑战性的工作。存储网络和基于对象的存储在这方面做了有益的探索,为统一虚拟存储结构的发展做了启发和技术准备。

本文探讨了网络存储的统一和虚拟化技术的现状和存在的问题,指出了统一化和虚拟化共同的终极发展目标是统一虚拟存储网络,通过统一虚拟存储体系结构将所有的存储资源统一起来,提供统一的资源共享服务。

1 NAS和SAN的融合

1.1 NAS和SAN的特点

SAN(Storage Area Network;存储区域网)是一种利用Fibre Channel等互联协议连接起来的可以在服务器和存储系统之间直接传送数据的存储网络系统,存储设备和应用服务器之间采用的是block I/O的方式进行数据交换。SAN具有高性能、高可用性、较高的灵活性、扩展方便、易于集中管理和可靠性高等优点,也存在设备造价高、需要专业维护人员、设备的互操作性较差等缺陷^[1]。

^{*})本文受国家自然科学基金项目(编号:60173043和60273073)和国家“973”重大基础项目(编号:2004CB318203)资助。张成峰 博士研究生,研究方向:基于网络的存储系统,海量存储。谢长生 教授,博士生导师,研究方向为:基于网络的存储系统,计算机高速接口与通道,采用新原理的超高密度、超高速存储技术。

NAS(Network Attached Storage:附网存储)是可以直接联到网络上向用户提供文件级服务的存储设备,具有可即插即用、总拥有成本(TCO)低的优点,但亦存在 I/O 速度不高、数据备份需要占用网络带宽等缺点^[2]。

显然,NAS 和 SAN 是两种截然不同的技术,尤其是两者构建技术不同,对企业来说,需要同时构建和管理两种不同的存储结构,增加了实现和管理成本,同时也不能很好地做到存储资源的整合。另一方面,NAS 和 SAN 又是相互补充的,两种结构可以因满足不同的需求而同时存在。

1.2 NAS 和 SAN 的融合方案

鉴于 NAS 和 SAN 的各自优点和缺点,要求 NAS 和 SAN 进行融合的呼声很高,也出现了一些相关产品,对已经产生的众多产品的方案进行分析,我们发现这些方案基本上可以分为两类:基于 NAS 和基于 SAN 的融合方案。

1.2.1 基于 NAS 的融合方案 该方案是将操作系统(文件系统)和存储子系统相互独立,其文件系统功能部分单独构成一个 NAS head,用以保存元数据,同时完成文件系统的功能;其存储子系统采用外置式的 SAN 结构。

该方案采用的是以 NAS 的方式和网络相连,因此可以实现异构的文件访问和共享;客户则是通过 IP 网络访问,可以充分利用现有的网络环境;存储子系统部分采用 SAN 结构,具有 SAN 高可扩展性、高可用性,以及存储资源的集中和统一管理等优点。

该方案的问题在于 NAS head:控制信息和数据信息都必须通过 NAS head,使得 NAS head 成为一个潜在的性能瓶颈并容易导致系统单点失效,为此需要采取冗余等措施加以解决。

1.2.2 基于 SAN 的融合方案 该方案的一个特点就是 把文件系统功能部分从应用服务器中剥离出来,集中到元数据服务器上,从而实现了命令通道与数据通道的分离,或者说是元数据与数据的分离存储;而应用服务器上都安装有一个专用的过滤程序,来处理元数据。

在应用服务器看来,访问方式采用的是文件 I/O 方式,整个存储系统相当于一个 NAS 文件服务器,因此具有 NAS 的优点;文件系统和存储子系统之间采用 SAN 的结构,因而该方案也具有 SAN 的优点。

整个系统只有一个元数据服务器,不过它只是处理元数据及控制信息,因此元数据服务器并不会是性能瓶颈。

1.2.3 统一存储网 统一存储网(USN; Unified Storage Network)则是由学术界提出的一个 NAS 和 SAN 融合方案,该方案综合了上述两方案的优缺点^[3]。

USN 采用 TCP/IP 作为存储网络的构建技术,通过 IP 互联设备将应用服务器、存储设备等互连成一个基于 IP 的 SAN,从体系结构的角度来说,具备了传统 SAN 高性能、易扩展等特点;同时,通过高速附网通道将存储子系统直接联入 LAN,从而形成了一个广义的 NAS。

和上述 NAS/SAN 融合方案相比,USN 具有如下特点:1、支持块设备和文件设备的同时接入,并进行统一管理,形成单一的命名空间。2、存储网络中,文件设备的引入,USN 可以同时提供块 I/O 服务和文件 I/O 服务。3、引入附网直连通道(NAC),不需要由应用服务器进一步处理的数据可经 NAC 直接返回,避免了应用服务器的存储转发,实现客户与存储设备之间的数据传送。

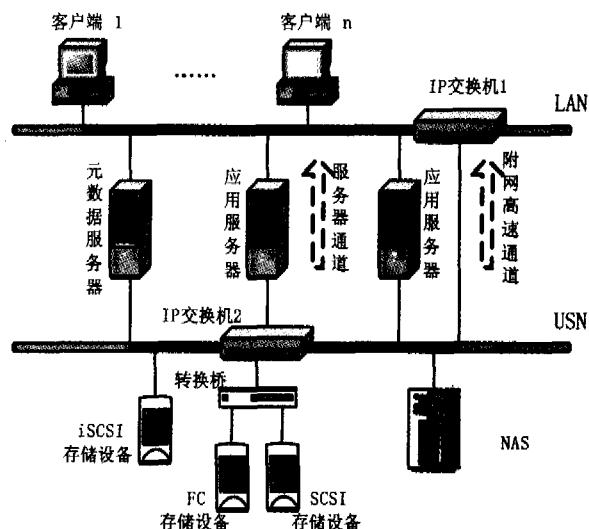


图 1 基于 TCP/IP 的统一存储网络系统(USN)

NAS 和 SAN 的融合方案很多,但是都不能同时完全发挥二者的优点,影响的因素很多,其中缺乏 NAS 和 SAN 进行综合管理的软件管理策略是一个重要原因,NAS 和 SAN 的融合,毕竟不是简单地将它们进行连接或重构。另一方面,新的存储技术和体系结构的不断出现:IP 存储、InfiniBand、DAFS 等,不但影响 NAS 和 SAN 本身的发展,也将会催生新的融合方案。另外,Ethernet 的发展经历了 10Mbps 到 100Mbps 再到 1000Mbps 的过程,现在,10Gbps 的 Ethernet 也即将问世,Ethernet 的发展对 NAS 和 SAN 的融合也必将带来很大的影响。

2 基于存储网络的虚拟化

存储虚拟化就是在存储系统中增加一个虚拟化层,通过虚拟化层,把存储系统中各种分散的、异构的存储设备映射成一个虚拟存储池。虚拟化层屏蔽了所有存储设备的物理特性,存储系统中所有存储设备对用户透明,用户则通过虚拟化层提供的接口,对虚拟存储池进行 I/O 访问。

基于存储网络的虚拟化的目标是,把分散在网络中的存储设备统一起来,使用户感觉和操作起来,像一个普通的存储设备,用户并不需要知道设备的供应商,也不需要知道设备的具体位置。

网络级的存储虚拟化主流是基于 SAN 的虚拟化。基于 SAN 的存储虚拟化,大多采用一个专用的元数据服务器,作为虚拟化层,这台专用的服务器称为元数据服务器或者元数据控制器。根据该服务器是否在数据传送的路径上,基于 SAN 的存储虚拟化的实现方式分为对称结构(Symmetric)和非对称结构(Asymmetric)两种实现模型^[4]。

2.1 对称方式的 SAN 存储虚拟化结构

对称结构也称为带内(in-band)模型,实现途径是在数据通路中插入虚拟存储层-元数据服务器,并由虚拟化管理软件对存储设备和服务器主机进行配置和管理,存储数据和控制信息使用同一条通路。

这种结构实现简单,而且将虚拟化层完全实现在专用虚拟平台上-元数据服务器,对应用服务器和存储设备的影响达到最小,因此对异构环境具有很强的适应性。

对称结构的不足主要在于,数据传输和元数据访问都必须经过元数据服务器,元数据服务器很可能成为系统瓶颈,制

约 SAN 的整体性能,并影响 SAN 的可扩展性。此外,这种结构的虚拟存储系统容易出现服务器到存储设备的单点故障,因此对称虚拟存储结构在实际使用中往往要做冗余配置。

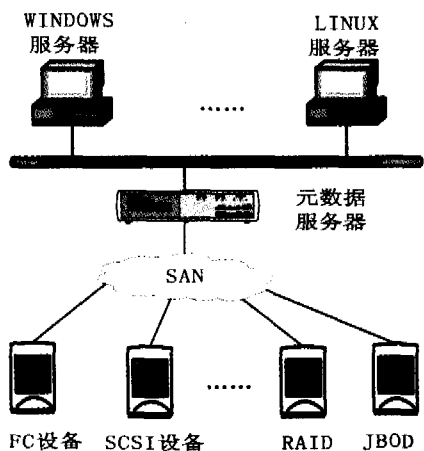


图 2 对称方式的 SAN 存储虚拟化结构

2.2 非对称方式的 SAN 存储虚拟结构

非对称结构也称为带外(out-of-band)模型,其实现途径是把装有虚拟化管理软件的主机或控制器从数据通路的外部接入存储网络,对网络进行管理。而网络存储系统除数据通路之外,还存在另外一条控制通路,存储数据和控制命令分别走这两条不同的通路。应用服务器节点的 I/O 请求先通过控制通路传送到专用的元数据服务器或控制器,获得元数据和数据视图后通过数据通路获得需要的数据。

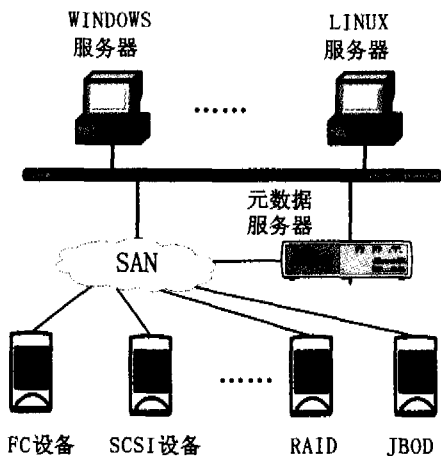


图 3 非对称方式的 SAN 存储虚拟化结构

在非对称性虚拟存储结构中,存储数据在专用的数据通路上传输,减少了网络延迟,增加了带宽的可用性,从而提高了系统性能,同时这种结构还避免了系统的单点故障和瓶颈。但是,由于虚拟化层的部分模块嵌入在应用服务器的操作系统中,造成实现工作相对比较复杂,对存储环境有依赖性,不适合有多种操作系统的异构存储环境,同时在一定程度上增加了用户的投资。

2.3 影响基于网络存储的虚拟化发展的因素

当前,影响存储虚拟化发展的最主要因素,就是存储虚拟化技术还缺乏行业标准。尤其是对于基于网络存储的虚拟化来说,存储用户、存储服务代理和存储资源三者之间非常需要开放性的标准。

由于存储虚拟化可以基于不同层次的方案或标准的软硬

件实现,这就意味着设备的选择非常广泛。迄今为止,许多厂商还忙于发展技术力量来创建各自的虚拟存储设备。由此可见,广大设备提供商和业内人士加强合作,尽快制定业界公认的标准,才能推动虚拟存储技术持续发展,为用户带来更多收益。

3 统一的虚拟存储网络

存储资源共享是网络存储体系结构发展的最本质的需求,网络存储的将来必然是统一虚拟存储,通过统一虚拟存储体系结构将所有的存储资源统一起来,提供统一的资源共享服务。

因此,我们可以把网络存储的统一化和存储虚拟化看作实现统一的虚拟存储网络的两个不同的途径,它们发展的最终目标是一致的,只是技术手段的不同。

另一方面,两者之间又是相互关联的,NAS 和 SAN 的融合过程,实质上也是对 NAS 和 SAN 的虚拟化过程。

我们以 USN 方案为例:

(1)当 USN 提供 block I/O 服务时,由于采用了附网直通通道(NAC)技术和三方传送协议,元数据和数据通道分离(数据流程见图 4),此时 USN 实际上相当于一种非对称的结构。

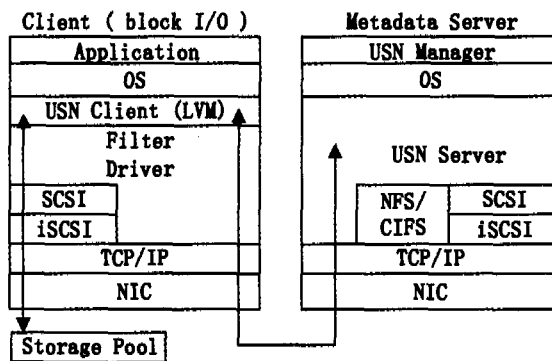


图 4 USN 提供 Block I/O 服务时的数据流程示意图

(2)当 USN 提供 file I/O 时,元信息、数据信息都必须通过元数据服务器(数据流程见图 5),此时 USN 实际上相当于一种对称的结构。

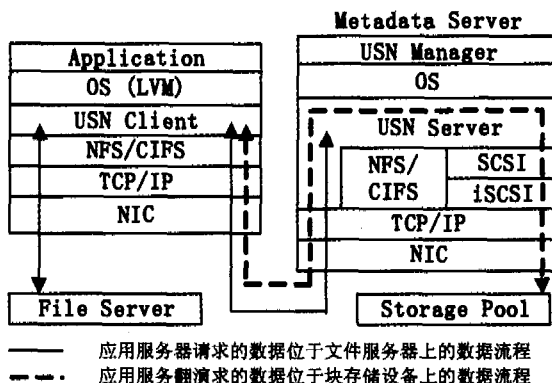


图 5 USN 提供 File I/O 服务时的数据流程示意图

目前,统一的虚拟存储网络还没有真正实现的方案,一些已经出现的方案,从原理上讲,只是对现有的存储虚拟化方案的扩展。以下两项技术的发展,将会为统一的虚拟存储网络提供启示和技术准备。

存储网络:存储网络实质上是网络技术和存储技术的结合,吸取了两者之间的优点,和普通的存储网络相比,具有可靠性高、管理灵活和可扩展性强等优点。存储网络还没有明确的标准出现,但是可以肯定的是,存储网络的实现,必须要求存储虚拟化的概念和技术出现突破,以便为用户提供更加优质的服务,在整合存储网络方面,网络技术和存储技术的结合是值得期待的^[7,8]。

基于对象的存储:基于对象的存储系统为每一个对象及其元数据分配一个独特的地址,不会出现重复,所有对象和元数据的存档和检索操作,都在一个巨大的平面地址空间内进行。在基于对象的存储系统中,客户端可以直接访问存储设备,减少了数据存储路径中的控制路径。基于对象的存储综合了NAS和SAN的优点,同时具有SAN的高速直接访问和NAS的数据共享等优势,提供了具有高性能、高可靠性、跨平台以及安全的数据共享的存储体系结构。基于对象的存储的标准最初来自于卡内基梅隆大学(Carnegie Mellon University, CMU)的1995~1999年间的“Network Attached Secure Disks”(NASD)项目。当前,对象存储设备(Object Storage Device)的标准正在制定之中^[6]。

4 基于自主式存储中介(ASAS)的网络存储系统

目前阻碍统一的虚拟存储网络的形成的主要因素是:

(1)缺乏统一的标准。除NAS和SAN外,网络存储技术还存在着其他繁多的标准和协议,而各种标准和协议都有自身的长处和不足,有效地整合不同存储网络将是极具挑战性的工作。

(2)存储管理日益复杂。存储网络使得存储容量的扩展性较好,但是带来管理的复杂性和成本呈指数增加。各种网络存储系统在体系结构上有很大的不同,面向的应用也各有差异,目前还缺乏有效存取模式去统一存取网络存储中的数据。进一步增加可共享的存储资源的数量变得难上加难。

(3)不能根据应用来提供相应的最佳存储服务。各种存储技术更多的是针对存储系统本身来提高对应用程序的服务性能,这种方式没有着重从应用程序的角度出发,根据应用程序的特点来动态地选择合适的存储技术和确定最优的存取访问策略。

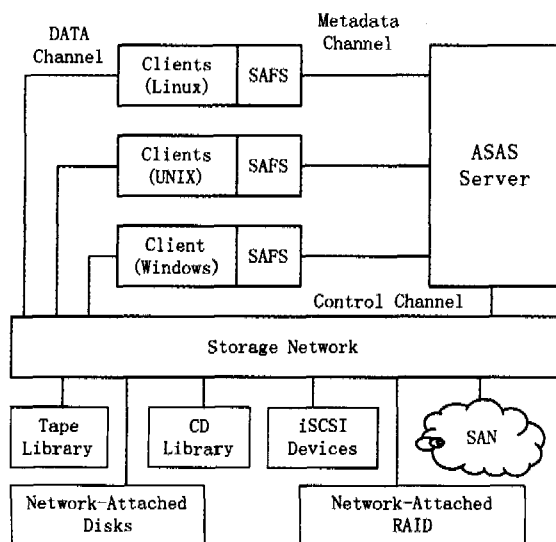


图6 基于自主式存储中介的存储网络

为了克服上述缺点,本文提出一种基于自主式存储中介

(ASAS, Autonomic Storage Agency System)的网络存储系统,该系统结合了网络存储和存储虚拟化技术,可以看作一个统一的虚拟化存储网络的雏形。

该系统以存储网络为基础,在应用程序与网络存储系统之间插入一个ASAS,它由ASAS服务器和驻留在各种应用主机中的SAFS(Storage Agency File System)两部分构成。系统运行过程中,各种存储资源(DAS、NAS、iSCSI等设备)将自动或交互地向ASAS服务器汇报当前存储系统属性、状态以及各种相应的参数,使得ASAS能够对整个存储系统进行管理和性能优化。应用主机通过SAFS提出的数据存储服务请求到达ASAS服务器后,ASAS服务器将根据应用主机的服务类型,安排合适的存储资源,确定最佳的存储策略,然后分别送到存储网络中相关的存储设备。当数据存取时,存储设备直接和应用主机的SAFS交互,从而降低ASAS服务器负载,缩短响应时间。另外,数据备份、灾难恢复、数据迁移等工作在ASAS服务器的监督管理下进行。

应用主机中的SAFS,即ASAS用户级文件系统,可以驻留在多种操作系统平台上(UNIX, Windows, Linux),为用户或应用程序提供对网络存储系统的数据访问服务。SAFS可以运用启发式方法获得用户对数据存储的具体需求(例如可用性、安全性、性能等要求),然后根据这些服务特征对数据进行合理的组织,尽可能地满足用户的这些要求。

结束语 随着网络技术的发展,信息量大幅增加,网络存储技术随之有了很大的发展。由于存在诸多的标准和协议,存储网络的管理复杂程度增加,不能充分利用现有的存储资源为用户提供更好的服务。

NAS和SAN的融合以及存储虚拟化技术从不同的角度、不同程度地解决了一些存在的问题,但是上述问题的彻底解决,还是需要新的体系结构的出现,比如统一虚拟存储体系结构。

统一的虚拟存储体系结构应该是从存储虚拟化的实质出发,无论从主机、网络还是存储设备看,存储资源就是一个单一存储映像,尽可能为用户提供优质的存储服务。统一的虚拟存储体系结构的完全实现还存在很多困难,不管NAS和SAN的融合方案、存储网络还是对象存储都只能是部分的甚至是一个方面的实现。

基于自主式存储中介(ASAS)的网络存储系统可以看作是统一的虚拟存储体系结构的一个实现方案的雏形。目前该方案的研究工作还在继续进行中。

参考文献

- [美]Farley M著,孙功星,等译. SAN存储区域网络(第2版). 机械工业出版社,2002. 12~34
- Gibson G A, Meter R V. Network attached storage architecture. Communications of the ACM,2002, 43(11): 37~45
- Fu Xianglin, Xie Changsheng, Liu Ruifang, Pan Shenggang. Performance Analysis of iSCSI-Based United Storage Network Merging NAS and SAN. Journal of Zhejiang University SCIENCE, 2004,5(1):1~7
- Milligan C, Selkirk S. Online storage virtualization; the key to managing the data explosion, System Sciences,2002. HICSS. In: Proc. of the 35th Annual Hawaii Intl. Conf. on 7~10 Jan. 2002. 3052~3060
- Baird R. Mass Storage Systems,1995. Virtual storage architecture guide(VSAG). In:Proc. of the Fourteenth IEEE Symposium on 11~14 Sept. 1995, 312~326
- Azagury A, Dreizin V, Factor M, et al. Towards an object store. Mass Storage Systems and Technologies, 2003. (MSST 2003). In: Proc. 20th IEEE/11th NASA Goddard Conf. on 7~10 April 2003. 165~176
- Network Appliance, Inc. Storage Grid Overview. <http://www.netapp.com/ftp/grid-overview.pdf>
- Hewlett-Packard Development Company, whitepaper, Storage Grid: The next step towards the Storage Utility