

GAOKEYONG CUNCHUWANGLUO
GUANJIANJISHU DE YANJIU

高可用存储网络 关键技术的研究

韩德志 傅 丰 著



科学出版社

www.sciencep.com

高可用存储网络关键技术的研究

韩德志 傅 丰 著

科 学 出 版 社

北 京

版权所有, 侵权必究

举报电话: 010-64030229; 010-64034315; 13501151303

内 容 简 介

本书在构建基于 TCP/IP 的统一存储网和海量固定内容网络存储系统的基础上, 围绕构建高可用存储网络的关键技术展开讨论。包括存储网络高可用概念和研究现状; 存储网络高可用系统设计的理论基础及关键技术; 面向恢复的自适应复制协议和面向恢复的容错算法的研究; 存储网络安全文件系统和安全算法的研究; 基于 TCP/IP 的统一存储网原型系统的研究与设计, 及基于内容的海量固定内容存储技术的研究; 存储网络高可用元数据服务器系统及统一存储网高可用系统的研究与设计; 总结并对后续研究工作进行展望。

本书对从事存储网络系统和高可用网络存储系统研究与设计的科技人员具有重要的参考价值, 还可作为计算机、信息技术等专业的研究生学习、研究的参考书。

图书在版编目(CIP)数据

高可用存储网络关键技术的研究/韩德志, 傅丰著. 北京: 科学出版社, 2009
ISBN 978-7-03-024950-0

I. 高… II. ①韩… ②傅… III. 计算机网络—信息存贮—研究 IV. TP393

中国版本图书馆 CIP 数据核字(2009)第 112733 号

责任编辑: 张颖兵 / 责任校对: 梅莹
责任印制: 彭超 / 封面设计: 苏波

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

武汉市科利德印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2009 年 7 月第 一 版 开本: A5(890×1240)
2009 年 7 月第一次印刷 印张: 9
印数: 1—1700 字数: 278 000

定价: 38.00 元

(如有印装质量问题, 我社负责调换)

前 言

网络应用的普及和数据信息爆炸性增长,使企业对存储网络系统的依赖性变得越来越强。一方面企业通过存储网络存放作为企业重要财富的数据信息,另一方面通过存储网络满足用户的多种需求和提供各种相应的服务,这些要求企业必须构建高可用的存储网络系统。目前,已经出现了多种网络存储技术,虽然这些技术构建的网络存储系统在一定程度上满足了应用对大容量、高性能、动态可扩展性、易维护性和开放性等方面的需求,但是在满足应用对高可靠性、高可用性的需求方面却显得不足。所以,存储网络的高可用技术的研究具有重要的理论意义和重大的经济价值。

构建高可用存储网络涉及多种技术,其关键技术主要包括:

(1) 冗余技术,包括各种设备冗余、网络路径冗余和相应的软件冗余等。利用冗余技术,可以保证在系统硬件失效、应用程序失效或系统软件失效时系统的可用性。

(2) 日志技术,记录存储网络系统接收和响应的 I/O 请求信息。当某一结点或程序出现故障时,通过保存的日志信息恢复替代结点的状态,以及应用程序失效前的状态,从而保证系统结构和数据的一致性,也保证用户 I/O 请求响应的连续性,日志技术往往同检查点和记录技术相结合。

(3) 复制技术,包括主动复制、被动复制、镜像和 RAID 等技术。通过这些技术,可以保证数据信息的可用性。

(4) 安全技术,通过防火墙和身份认证阻止非法用户对存储网络系统的访问,防止黑客对系统造成各种破坏以及大量非法访问造成系统拥塞,从而保证有权限用户的正常使用。通过访问控制来控制存储网络中的某些卷或者目录,甚至一些文件,使一些用户可以访问它们,其他用户则不能访问,从而保证用户在存储网络中的个人信息的安全性和可用性。

(5) 负载均衡技术,通过 I/O 调度和数据迁移技术使存储网络中的各个设备(应用服务器、元数据服务器及网络设备)、存储结点等都能很好

地工作,从而保证用户的 I/O 请求信息得到快速和高效的响应。

(6) 系统失效监测处理技术,通过失效监测技术,可以即时发现系统的故障设备、结点或应用程序,然后通过替换和状态恢复或重启,来保证系统服务的连续性。

本书在构建基于 TCP/IP 的统一存储网和海量固定内容网络存储系统的基础上,围绕构建高可用存储网络的关键技术展开讨论。全书共分 9 章,第 1 章介绍存储网络高可用技术相关的概念和研究现状;第 2 章讨论了存储网络高可用系统设计的理论及关键技术;第 3 章针对传统的复制协议存在的缺陷,研究了存储网络面向恢复的自适应复制协议和面向恢复的容错算法;第 4 章在讨论复杂存储网络安全技术的基础上,重点研究了存储网络安全文件系统和安全算法;第 5 章分析了传统网络存储技术,研究并设计了基于 TCP/IP 的统一存储网,这是存储网络高可用系统研究的原型系统,是第 6~8 章研究的基础;第 6 章介绍了固定内容存储技术,研究了基于内容海量固定内容存储技术;第 7~8 章针对统一存储网原型系统,研究并设计了存储网络高可用元数据服务器系统和统一存储网络高可用系统;第 9 章对书中提到的主要研究工作进行总结,并展望了后续研究工作。

本书是作者相关研究工作的总结。这些研究工作是国家自然科学基金项目“统一存储网(USN)理论、结构和实验研究”(60173043),国家重点基础研究发展计划 973 项目“海量数据网络存储系统的研究”,河南省科技计划资助项目“电子政务内网数据安全系统的研究与设计”(072100451230),国家高技术研究发展计划(863,2007AA01Z449)“基于应用行为规范的网路主动实时防护系统”的研究内容。本书是由广东外语外贸大学学术著作出版基金和河南省科技计划资助项目(072100451230)资助出版,并得到广东外语外贸大学信息学院和科学出版社的大力支持,在此一并表示衷心感谢。

限于作者学识水平,疏漏之处在所难免,敬请同行和读者批评指正。

韩德志 傅 丰

2009 年 1 月

目 录

前言	1
第 1 章 存储网络的高可用技术概述	1
1.1 研究背景	1
1.2 存储网络高可用实现技术概述	14
1.3 存储网络高可用技术研究现状	24
1.4 存储网络高可用技术研究的意义	32
第 2 章 高可用系统设计理论及关键技术	34
2.1 存储网络及高可用系统涉及的概念	34
2.2 存储网络高可用系统设计的理论分析	36
2.3 存储网络高可用系统设计的关键技术概述	48
2.4 存储网络高可用技术面临的新挑战	60
第 3 章 面向恢复的自适应容错技术研究	62
3.1 研究背景	63
3.2 面向恢复的自适应复制协议研究	64
3.3 面向恢复的自适应容错算法研究	75
3.4 本章小结	83
第 4 章 存储网络安全问题的研究	84
4.1 统一存储网络(USN)安全问题的研究	84
4.2 MUVFS 安全系统设计	86
4.3 基于 USN 的安全算法研究	96
4.4 安全系统对 USN 性能的影响	100
4.5 本章小结	103
第 5 章 统一存储网络的研究	104
5.1 USN 体系结构	104
5.2 USN 互连技术	112
5.3 USN 单一命名空间的实现	122

5.4	USN 的数据备份系统的设计与实现	132
5.5	USN 实验评估及性能分析	137
5.6	本章小结	144
第 6 章	海量固定内容存储技术的研究	145
6.1	研究背景	145
6.2	内容存储与内容管理的融合	148
6.3	固定内容存储所涉及的关键技术	152
6.4	MFCSS 元数据管理系统	166
6.5	MFCSS 系统性能测试与评价	197
6.6	MFCSS 与 FTP 性能比较	204
6.7	本章小结	207
第 7 章	双机高可用系统的设计方法与性能分析	209
7.1	研究背景	209
7.2	双机高可用系统的实现方案	212
7.3	实现双机高可用系统关键问题	221
7.4	双机高可用系统性能分析	222
7.5	本章小结	228
第 8 章	存储网络高可用系统的研究与设计	230
8.1	USN 系统概述	230
8.2	USN_HA 系统的设计思想	236
8.3	USN_HA 系统监测机制	239
8.4	USN_HA 的日志机制	241
8.5	USN_HA 冗余替换机制	244
8.6	USN_HA 失效恢复机制	250
8.7	USN 系统的可用性分析	253
8.8	本章小节	262
第 9 章	总结与展望	264
9.1	主要工作总结	264
9.2	进一步的研究工作构想	266
参考文献	268

第 1 章 存储网络的高可用技术概述

1.1 研究背景

随着互联网和电子商务的快速发展,网络数据信息呈现爆炸性增长趋势,存储网络技术越来越被企业所接受,并在企业的商业活动中充当非常重要的角色。企业依靠存储网络存储关键的数据和各种有用的信息,这些数据和信息是企业的重要财富,它们对企业的生存和发展起着至关重要的作用。用户可以从企业的存储网络中获得企业的产品信息,企业可以借助存储网络进行各种商业活动。构建高可用性的存储网络是至关重要的,若存储网络系统发生故障,如服务器故障或各种设备故障,造成系统服务中断将会造成客户的流失和企业收入的损失。研究表明,在网络存储系统中,若系统发生 1 小时的停机,企业将损失 15 万美元到 645 万美元之间^[1]。为了保证企业达到 24 小时×365 天的应用需求,必须构建高可用性的存储网络。高可用存储网络的重要性可从美国 911 事件得到印证。在 911 事件中,总部设在世贸中心的摩根-斯坦利公司,由于建有高可用的存储网络系统,在灾难发生的第二天就恢复了正常运转,而其他很多没有建立容灾备份高可用存储网络系统的公司,由于关键经营数据被毁灭,从而接近崩溃的边缘。表 1.1 给出了来自美国商业的有关不同类型企业因为宕机而带来的经济损失的评估报告^[2]。

表 1.1 宕机时间对企业的损失

企业部门	业务操作	企业损失/ (美元/小时宕机时间)	平均损失/ (美元/小时宕机时间)
金融	经济业操作	5 600 000~7 300 000	6 500 000
金融	信用卡/销售论证	2 200 000~3 100 000	2 600 000
媒体	视频点播	67 000~233 000	150 000
零售	电视购物	87 000~140 000	113 000

续表

企业部门	业务操作	企业损失/ (美元/小时宕机时间)	平均损失/ (美元/小时宕机时间)
零售	家庭网上销售	60 000~120 000	90 000
运输	航班预定	67 000~112 000	90 000
媒体	远程订票	56 000~82 000	69 000
运输	包裹运输	24 000~32 000	28 000
金融	ATM 费用	120 000~17 000	15 000

此外,国家宇航、军事等领域也迫切需要高可用的存储网络。在航空航天领域,信息存储系统要存放各种测试数据、计算数据以及卫星发回的各种观测数据,这些数据信息,如气象卫星和资源卫星发回的数据信息,对科学研究和人们的生活有着重要的影响,如果其存储网络被毁坏将造成不可估量的损失。在现代战争中,信息已成为一种重要的甚至起决定性的因素。在瞬息万变的战场,情报的获取对敌我双方胜负攸关。为了能得到战争的主动权,做到知己知彼,位于战斗前沿的作战单位、情报部门或军事侦察卫星等信息收集单元,将战斗的进行情况和敌方的战斗部署信息发送给指挥部,指挥部对所有的情报进行整理分析并作出相应的战略、战役决策。因此,存放有重要军事信息及关键性决策的军事存储网络便成为敌方攻击的主要目标,这对军用存储网络的高可用性提出了更高的要求。

1.1.1 网络存储技术的发展

用户把信息通过网络(包括 LAN、WAN 和 Internet)存储到连接在网络上的存储设备,或通过网络从连接在网络上的存储设备中获得所需信息的过程,叫做网络存储。我们将由网络存储设备、网络设备(如网络服务器、网络交换机、Hub、路由等)通过相关的协议(FC 协议、TCP/IP 协议,以及 iSCSI、FCIP、iFCP、Infiniband、mFCP 等)连接起来构成的网络,叫做存储网络。存储设备与网络连接,有直接连接和间接连接两种方式。直接连接是指存储设备通过网线直接连接到网络上,有权限的用户

可以直接访问其中的共享信息;间接连接是指存储设备通过其他设备连接到网络,如通过 NAS 头、应用服务器等。

按照存储设备与网络的连接方式,主流的网络信息存储系统有直连存储(direct attached storage, DAS)、附网存储(network attached storage, NAS)和存储局域网(storage area network, SAN)三种形式。传统存储结构多采用 DAS,现代网络应用的快速发展对传统存储结构提出了极大的挑战,促进了以 NAS 和 SAN 为代表的网络存储技术的成熟和快速普及。除此之外,还出现了多种新的网络存储技术,如基于 IP 的 SAN、对象存储技术等。

1. 直连存储

DAS 是一种以服务器为中心的存储结构,各种存储设备通过 IDE 或 SCSI 等 I/O 总线与服务器相连。所有的客户端请求与数据传送都经过服务器,数据的读写流程如下:客户发送请求给文件服务器;文件服务器解析该请求并将其传送到存储设备;存储设备将数据读出,并返回文件服务器,在文件服务器内部依次经过硬盘控制器、系统内存、网络接口卡等;文件服务器将数据传送给客户。由于存储系统附属于服务器,受服务器总线技术的限制,DAS 的可扩展性差。并且当客户连接数增多时,服务器将成为整个系统的性能瓶颈。所以,DAS 存储方式难以满足现代社会对网络存储系统大容量、高性能、动态可扩展等方面要求,解决这一问题的有效方法是将访问模式从以服务器为中心转化为以设备和网络为中心,这导致了网络存储技术(NAS 和 SAN)的普及和发展。

2. 附网存储

NAS 是一种以设备为中心的存储结构^[3-6]。按照存储网络工业协会(SNIA)的定义,NAS 是可以直接连接到网络向用户提供文件级服务的存储设备。作为一种存储设备,NAS 具有简化的实时操作系统,它可以将硬件和软件有机地集成在一起,用以提供文件服务。目前采用的协议是 NFS 和 CIFS^[7,8],其中 NFS 应用在 Unix 环境下,最早由 SUN 开发,而 CIFS 应用在 NT/Windows 环境下,是由 Microsoft 开发。NAS 的结构及采用的协议使得 NAS 具有以下优点:① 异构平台下的文件共享,不

同操作系统平台下的多个客户端可以很方便地共享 NAS 中的同一个文件;② 充分利用现有的 LAN 网络结构,保护现有投资;③ 容易安装,使用和管理都很方便,实现即插即用;④ 广泛的适用性,由于基于 IP/Ethernet 以及标准的 NFS 和 CIFS, NAS 可以适应复杂的网络环境;⑤ 广泛的适用性。另外,可以用 NAS 设备后端连接高性能的 RAID 或者 SAN 网络,使用户以 NAS 方式访问 RAID 和 SAN。

实际应用中 NAS 也表现出一些缺陷^[9,10]。首先,在文件访问的速度方面,由于 NAS 采用的是 file I/O 方式,在客户端 file I/O 请求要经过整个 TCP/IP 协议栈封装后经过网络传输。被封装的 file I/O 请求到达 NAS 后,同样要经过整个 TCP/IP 协议栈解封装再传到 NAS 的文件系统,最后对存储设备进行读写。数据从存储设备取出之后要经过类似处理 file I/O 请求相反的过程,这带来巨大的网络协议开销,这种 file I/O 的速度和 SAN 的 block I/O 相比很低。因此, NAS 不适合在对访问速度要求很高的应用场合,如数据库应用,在线事务处理。

其次,在数据备份方面, NAS 需要占用 LAN 的带宽,浪费宝贵的网络资源,严重时甚至影响客户应用的顺利进行。

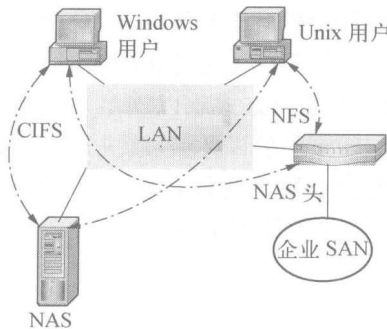


图 1.1 NAS 的拓扑结构

第三,在资源的整合和 NAS 的管理方面, NAS 只能对单个存储(单个 NAS 内部)设备之中的磁盘进行资源的整合,目前还无法跨越不同的 NAS 设备,难以将多个 NAS 设备整合成一个统一的存储池,因而难以对多个 NAS 设备进行统一的集中管理,只能进行单独管理。NAS 的拓扑结构如图 1.1 所示。

3. 存储局域网

SAN 是一种以网络为中心的存储结构^[11,12]。按照 SNIA 定义, SAN 是一种利用 Fibre Channel 等互连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。SAN 是一种体系结构,它是采

用独特的技术(如 FC)构建的、与原有 LAN 网络不同的一个专用的存储网络,存储设备和 SAN 中的应用服务器之间采用的是 block I/O 的方式进行数据交换。SAN 的拓扑结构如图 1.2 所示。

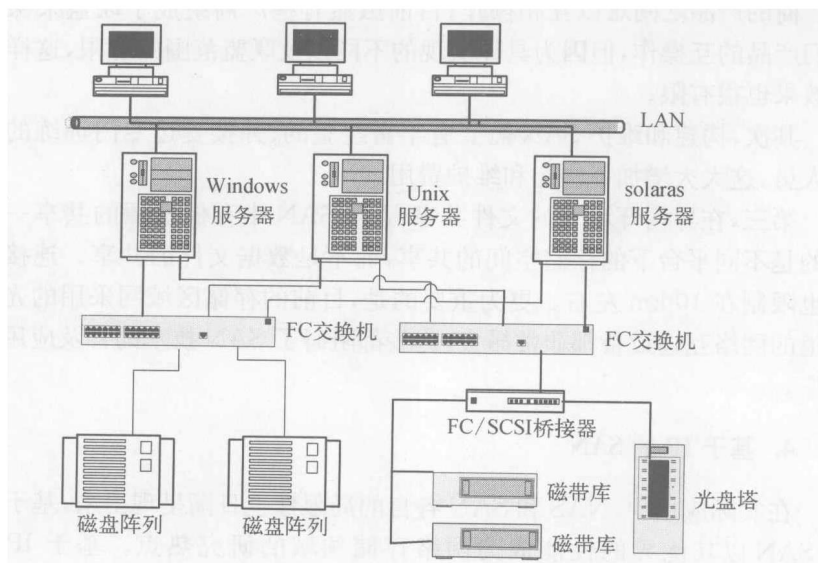


图 1.2 SAN 的拓扑结构

独特的体系结构和构建技术使得 SAN 具有如下优点^[9-11]：① 高性能、存取速度快，目前光纤通道可提供 4 Gb/s 的带宽，新的 10 Gb/s 的标准也正在制定之中；② 高可用性，网络用户可以通过不止一台服务器访问存储设备，当一台服务器出现故障时，其他服务器可以接管故障服务器的任务；③ 集中存储和管理，通过整合各种不同的存储设备形成一个统一的存储池，向用户提供服务，可以很容易地扩充存储容量；④ 高可扩展性，服务器和存储设备相分离，两者的扩展可以独立进行；⑤ 支持大量的设备，理论上具有 1500 万个地址；⑥ 实现 LAN-free backup，数据备份不占用 LAN 带宽；⑦ 支持更远的距离，另外通过光纤通道网卡、集线器、交换机等互连设备，用户可根据需要灵活地放置服务器和存储设备。在数据库等其他面向事务处理的应用场合，SAN 的优势更加明显。

具体应用中, SAN 有着自身的一些缺陷^[9-11]。首先, 设备的互操作性较差。目前采用最早和最多的 SAN 互连技术还是 Fibre Channel, 对于不同的制造商, 光纤通道协议的具体实现是不同的, 这在客观上造成不同厂商的产品之间难以互相操作; 目前虽然有些厂商组成了联盟来保证他们产品的互操作, 但因为具体实现的不同以及联盟范围的有限, 这样做的效果也很有限。

其次, 构建和维护 SAN 需要有丰富经验的、并接受过专门训练的专业人员, 这大大增加了构建和维护费用。

第三, 在异构环境下的文件共享方面, SAN 中存储资源的共享一般指的是不同平台下的存储空间的共享, 而不是数据文件的共享。连接距离也限制在 10 km 左右。更为重要的是, 目前的存储区域网采用的光纤通道的网络互连设备都非常昂贵, 这些都阻碍了 SAN 技术的普及应用和推广。

4. 基于 IP 的 SAN

在实际应用中, NAS 和 SAN 各自的局限性也日渐呈现出来, 基于 IP 的 SAN 以其优异的性能成为网络存储领域的研究热点。基于 IP 的 SAN 互连技术主要包括 FCIP, iFCP, iSCSI, Infiniband, mFCP^[12]。

FCIP 技术原理是将 FC 帧封装到 IP 数据包中, 再通过 IP 网络传输到另外一个 FC 的 SAN。目的 SAN 接受到这个 IP 包后, 将其解包使其恢复成封装之前的 FC 帧。通过 FCIP 协议可以方便地实现两个距离较远的 SAN 在 Internet 网络上相互通信。

iFCP 是将 FC 协议映射到 TCP 协议之上, 和 FCIP 不同, FCIP 是一种隧道(tunneling)协议, 除了将 FC 帧封装成 IP 数据包以外, 不对 FC 帧作任何处理; 而 iFCP 则是一种网关协议, 它对 FC 帧进行协议转换, 重新用 TCP/IP 协议来表达 FC 帧。它必须对 FC 帧进行更多的处理, 如读取 FC 帧的头部, 理解其地址, 并用 IP 地址方式来表示。

iSCSI 技术原理是将 SCSI 协议映射到 TCP/IP 之上, 即将主机的 SCSI 命令封装成 TCP/IP 数据包, 在 IP 网络上传输, 到达目的结点后, 再恢复成封装前的 SCSI 命令, 从而实现 SCSI 命令在 IP 网络上的直接、透明传输。

Infiniband 是一种可简化和加快服务器之间的连接,以及服务器与其他相关系统(诸如远程存储和网络设备)之间连接的全新输入/输出(I/O)技术。它通过一套中心机构(中心 InfiniBand 交换机)在远程存储器、网络以及服务器等设备之间建立一个单一的连接链路,并由中心 InfiniBand 交换机来控制流量。它的结构设计得非常紧密,大大提高了系统的性能、可靠性和有效性,能缓解各硬件设备之间的数据流量拥塞。Infiniband 技术目前主要被较大的数据中心所采用。mFCP(metropolitan fibre channel protocol),除了用 UDP 代替 TCP 外,mFCP 和 iFCP 是相同的。

目前,技术比较成熟并且被国际标准化组织推广为国际标准的是 iSCSI 协议^[12-14]。iSCSI 是一种融合 NAS 和 SAN 的技术,最初由 Cisco 和 IBM 两家发起,由 Internet 工程任务组(IETF)将其标准化。目前由 IETF 制定的 iSCSI 协议草案的最高版本为 draft-ietf-ips-iscsi-20。客户端要通过 iSCSI 协议同存储设备之间传输数据,客户端必须装载 iSCSI 客户端程序,即客户端作为 initiator,而存储设备端必须装载 iSCSI 服务器端程序,即服务器端作为 target。在工业上 iSCSI 的实现有两种方式:客户端/嵌入式存储设备方式;iSCSI Gateway 方式,即 IP/FC 桥接器方式。如图 1.3 所示,其工作流程为 iSCSI 客户端程序(initiator)包装 SCSI 命令,然后发送到 IP 网络上;iSCSI 目的端程序(target)从 IP 网络上接收 SCSI 命令;SCSI 命令要么经过嵌入式的 iSCSI 驱动程序传到存储设备,要么经过 iSCSI 网关传入 FC SAN。显然 IP Storage 的出现极大扩展了存储系统的范围,但另一方面,由于经过多次的协议转换以及现有传输协议不适应存储的特性,使得点对点的数据库操作和传输变得复杂而且安全性受到极大的挑战。

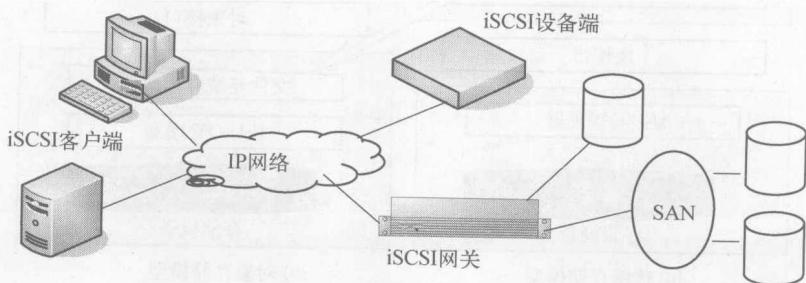


图 1.3 iSCSI 的实现方式

目前 iSCSI 配置有三种方式。第一种配置存储设备是一个单个磁盘,客户端通过 IP 网络直接访问存储设备;第二种配置存储设备是 SAN,客户端通过 IP 网络和 iSCSI 网关访问 SAN 中的存储设备;第三种配置同第二种类似,只是 iSCSI 网关中融合有 NAS 网关的成分,客户端可以 NAS 方式访问 SAN,也可以 iSCSI 的方式访问 SAN。

5. 对象存储技术

一个存储对象是存储设备上多个字节的逻辑集合,它包括访问数据的属性、属性描述、数据特征和阻止非授权用户访问的安全策略等^[15-19]。对象的大小可以变化,它可以存放整个数据结构,如文件、数据库表、医学图像或多媒体数据等。存储对象具有文件和块二者的优点:像数据块一样在存储设备上被直接访问;通过一个对象接口,能像文件一样,在不同操作系统平台上实现数据共享。

对象存储的结构如图 1.4 所示。它将文件系统分成两部分,一部分是与用户有关的部件,一部分是与存储有关的部件。与用户有关的部件放在文件服务器上,而与存储有关的部件则下放到存储设备上,在文件服务器同存储设备之间插入一个专用的对象接口,而不是块接口。由对象

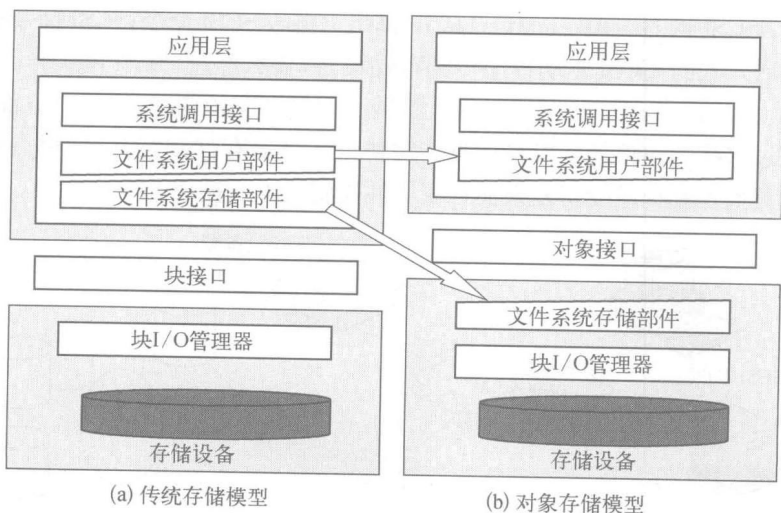


图 1.4 传统的模型与对象存储模型比较

存储设备构成的存储区域网如图 1.5 所示,客户端与文件管理服务器共享元数据,文件管理服务器负责对对象存储设备的管理,客户可以用与文件管理服务器共享的元数据直接对对象存储设备进行访问。对象存储结构的特性使其在网络数据存储环境中具有较大优势,主要体现在以下几方面。

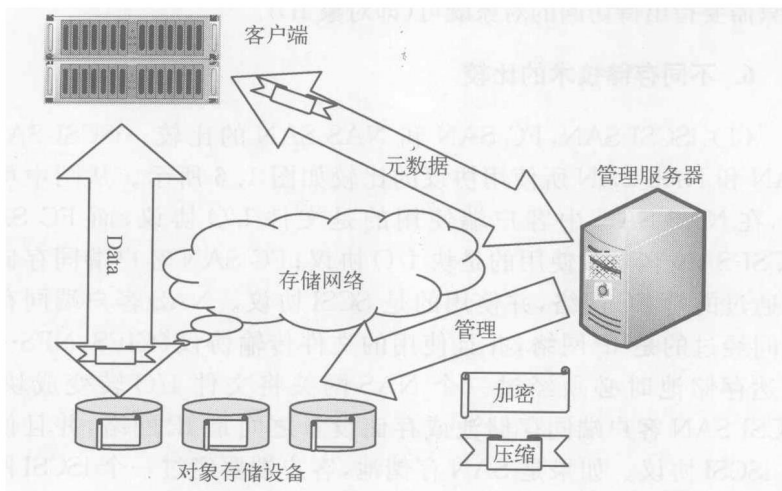


图 1.5 由对象存储设备构成的存储区域网的访问模式

(1) 性能优势。存储体系结构的最大竞争力在于存储性能。对象存储由于其存储设备的智能化以及访问的并行化,可以最大限度地利用网络的带宽,提高应用的存储能力。

(2) 存储设备的智能化。对象的自我管理功能大大增强了存储设备的智能化,降低了整个系统的管理负担,使其更高效。

(3) 数据的共享更容易。在对象存储系统中,任何客户端都可以通过访问提供的标准文件接口,访问存储设备上的数据,其 Cache 策略保证了客户端访问数据的一致性,使得文件级的数据共享更容易。

(4) 管理更方便。OSD 的自我管理减少了系统管理员的工作量,提高了管理质量。管理员只需要关注存储系统的逻辑视图,不需要知道设备的细节。此外,OSD 的智能化使其可对应用访问模式进行统计,并根据统计结果更有效地组织数据,提高设备的空间利用效率和访问效率。

(5) 更好的安全性。同块设备不同,对象可利用自身的属性对用户

访问进行验证,只有通过验证的请求才能执行 I/O 操作。

在基于对象的存储系统中,存储空间由存储设备管理和分配,而不再信赖运行在主机上的文件系统。主机系统中原先由应用程序执行的定位逻辑和由文件系统执行的存储空间管理功能迁移到存储设备中,应用程序只需要指出待访问的对象既可(即对象 ID)。

6. 不同存储技术的比较

(1) iSCSI SAN, FC SAN 和 NAS SAN 的比较。iSCSI SAN, FC SAN 和 NAS SAN 所使用协议的比较如图 1.6 所示。从图中可以看出,在 NAS SAN 中客户端使用的是文件 I/O 协议,而 FC SAN 和 iSCSI SAN 客户端使用的是块 I/O 协议;FC SAN 客户端同存储池之间通过的是 FC 网络,并使用的是 SCSI 协议。NAS 客户端同存储池之间经过的是 IP 网络,并且使用的文件传输协议(CIFS, NFS……),到达存储池时必须经过一个 NAS 网关将文件 I/O 转变成块 I/O。iSCSI SAN 客户端同存储池或存储设备之间是 IP 网络,并且使用的是 iSCSI 协议。如果是 SAN 存储池,客户端要经过一个 iSCSI 网关访问 SAN 存储池。

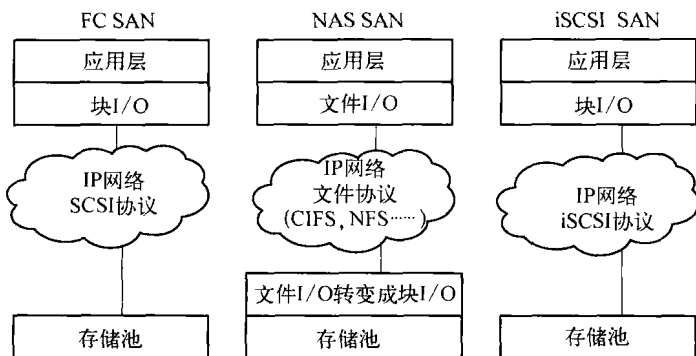


图 1.6 NAS SAN, FC SAN 和 iSCSI 所使用协议比较

(2) DAS, SAN, NAS 和 iSCSI 应用数据传输的比较。DAS, NAS, SAN 和 iSCSI 应用数据传输的比较如图 1.7 所示。从图中可以看出,在 DAS 中块 I/O 命令直接由主机的系统总线经 SCSI 总线适配器到达存储