

第 5 章 网络附接存储

5.1 NAS 概述

网络附接存储(network attached storage,NAS)是网络存储基于标准网络协议实现数据传输,是一种文件共享服务,拥有自己的文件系统,通过 NFS、SMB、CIFS、FTP 等协议对外提供文件级共享和数据备份。第一代 NAS 产品于 1999 年在市场上出现,由于当时 NAS 技术相对不够成熟,产品的功能比较有限,性能也不够稳定。2000 年之后,第二代 NAS 产品开始出现。由于广泛采用了 RAID、磁盘镜像、双机热备等新型技术,同时支持 HTTP、FTP、SMB、AFP、CIFS 等多种协议,并增强网络的安全认证机制,使得产品的性能得到了大幅提升。

NAS 包括硬盘阵列、DVD 驱动器、磁带驱动器或可移动硬盘等存储器件以及专用服务器。安装有专门操作系统的专用服务器,可充当远程文件服务器,利用 NFS、SMB/CIFS、FTP 等协议对外提供文件级的访问服务,典型拓扑如图 5-1 所示。

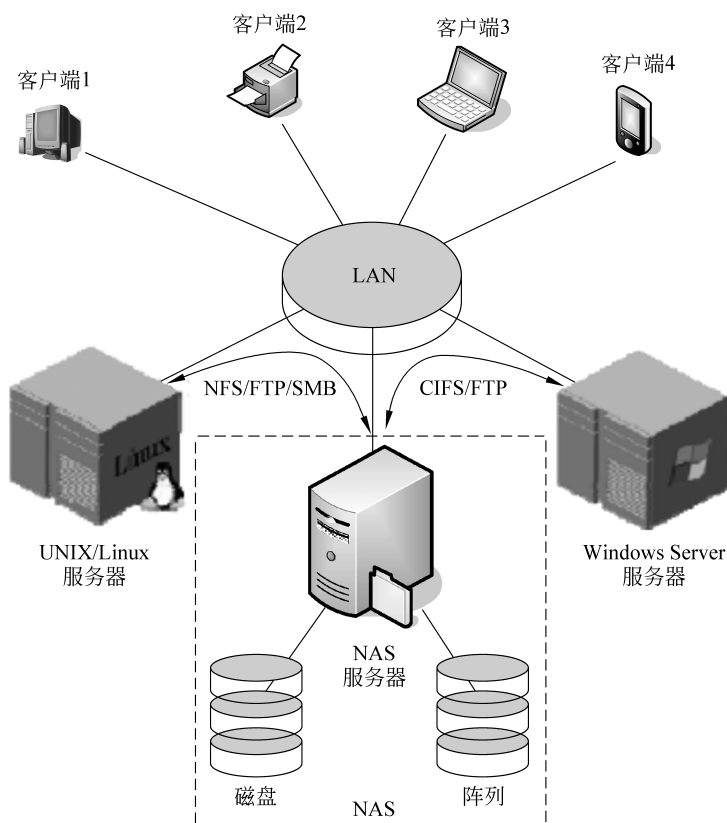


图 5-1 NAS 网络拓扑图

NAS 是一种利用可直接与网络介质相连的特殊设备实现数据存储的机制。由于这些设备都分配有 IP 地址,所以客户机通过充当数据网关的服务器可以对其进行存取访问,在某些情况下,甚至不需要任何中介介质客户机也可以对其进行直接访问。

NAS 设备可实现集中存储与备份,具有如下主要特点。

(1) 提供了一个高效、低成本的资源应用系统。由于 NAS 本身就是一套独立的网络服务器,可以使 NAS 主机、客户机和其他设备广泛分布在整个企业的网络环境中,提供可靠的文件级数据整合。可以基于已有的企业网络方便接入,具有非常好的可扩展性。

(2) 适用于需要通过网络将文件数据传送到多台客户机上的用户。在数据必须长距离传送的环境中,NAS 设备可以很好地发挥作用。

(3) 应用于高效的文件共享任务中,例如 UNIX 中的 NFS 和 Windows 中的 CIFS,其中基于网络的文件级锁定提供了高级并发访问保护的功能。采用 NFS、SMB/CIFS 提供异构网络之间的文件共享,屏蔽了操作系统和文件系统的差异。

(4) 提供灵活的个人磁盘空间服务。NAS 可以为每个用户创建个人的磁盘使用空间,方便用户查找和修改自己创建的数据资料。

(5) 有效保护资源数据。NAS 具有自动日志功能,可自动记录所有用户的访问信息。嵌入式的操作管理系统能够保证系统永不崩溃,以保证连续的资源服务,并有效保护资源数据的安全。

(6) 实现了多服务器之间的存储共享。文件共享(即文件服务器)是 NAS 最基本的应用。文件服务器通过一个文件系统管理磁盘阵列,存取磁盘阵列上的文件,并管理相应的网络安全和访问授权。

(7) 提供数据在线备份的环境。NAS 支持外接的磁带机,能有效地将数据从服务器中传送到外挂的磁带机上,保证数据安全、快捷备份。

(8) 备份和容灾。NAS 网络存储器的另一项重要功能是备份和容灾,大多数 NAS 都具有多种备份功能,包括本地备份(将计算机上的数据通过局域网备份到 NAS 中)、异地备份(将异地计算机上的数据通过广域网备份到 NAS 中)和 NAS 间备份(NAS 与 NAS 之间复制数据)等。部分 NAS 还具有一键备份功能,可将闪存和外置硬盘等 USB 存储设备插入 NAS 的特定 USB 接口,通过备份按钮就能把 USB 存储设备上的文件备份到 NAS 中。此外,具有两个硬盘位的 NAS 可以组建 RAID 0 和 RAID 1 系统,其中 RAID 0 系统具有较好的磁盘性能,RAID 1 系统具有较好的安全性。具有 4 个硬盘位的 NAS 则可以组建更高级的 RAID 5 系统,在保障数据安全的同时还能提高磁盘性能。

NAS 可以利用现有的 IP 网络设施,以网络共享的方式(如 CIFS 和 NFS)为终端客户机提供一致的文件系统,所以 NAS 存储解决方案在数据中心是一个极佳的选择。伴随着万兆以太网的出现和投入商用,存储网络带宽大大提高了 NAS 存储的性能。NAS 摆脱了服务器和异构化架构的桎梏,在解决了足够的存储和扩展空间的同时,还提供极高的性价比。

5.2 NAS 的存储架构

5.2.1 整体结构

NAS 包括核心处理器,文件服务管理工具,一个或者多个的硬盘驱动器。NAS 强调的是网络和存储功能,所以在构造硬件结构中只考虑网络和存储两方面的问题。

NAS 在一个 LAN 中占有自己的结点。NAS 设备包括存储器件和集成在一起的简易服务器(功能服务器),可用于实现涉及文件存取及管理的所有功能。NAS 基本硬件结构如图 5-2 所示。

存储子系统主要是提供对 IDE、SCSI、SATA 技术的支持,提供工业标准 EIDE 控制器、SCSI 控制

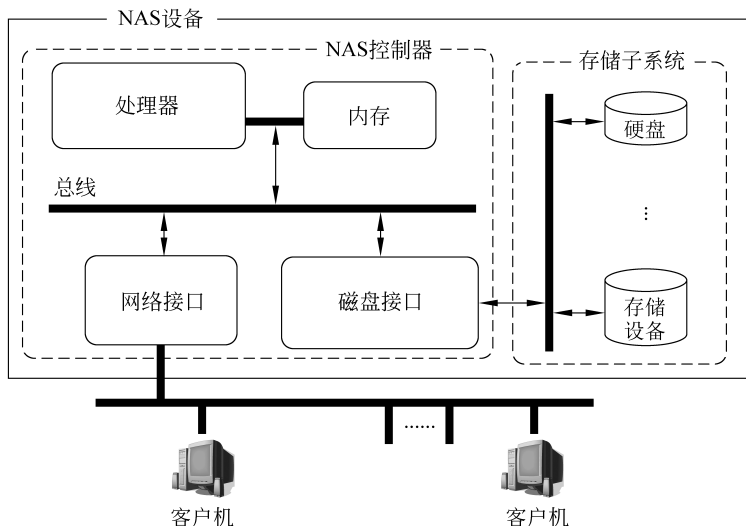


图 5-2 NAS 基本硬件结构

器、阵列控制器,使得系统可以连接光碟塔、磁盘阵列等各种设备。

网络控制模块用于实现网络适配器的功能,主要进行数据帧的生成、识别与传输、数据编译、地址译码、数据传输的出错检测和硬件故障的检测等。它最终提供一个普通网络接口和高速光纤通道连接口,传输速率不低于 100MB/s,可使系统方便地与以太网相连或挂接在由高速光纤通道连成的 SAN 上。系统的核心操作系统和相关系统软件都可以固化在 ROM 或者闪存上,供系统启动时读取。

由于去掉了系统中不必要的系统控制以及处理功能,从而减少了系统的成本,提高了存储系统的吞吐率和通信能力。

5.2.2 NAS 软件构成

NAS 系统软件可划分为五大模块:操作系统、卷管理器、文件系统、网络文件共享和 Web 管理,如图 5-3 所示。

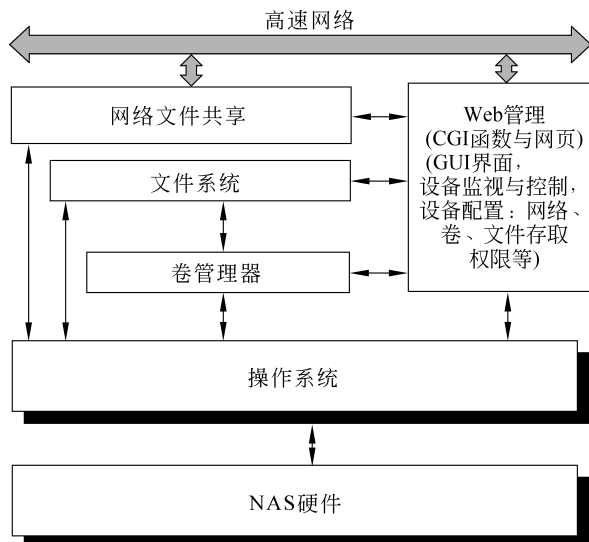


图 5-3 网络存储服务器软件组成

1. 操作系统

NAS 不依赖通用的操作系统,而是采用一个面向用户的、专门用于数据存储的简化操作系统,由于其中内置了与网络连接所需的协议,所以可使整个系统的管理和设置较为简单。NAS 操作系统通常需要实现 4 方面的功能:设备驱动功能、设备管理功能、文件共享服务功能和应用系统功能。NAS 操作系统是为文件做了优化的,具有多线程、多任务的高稳定内核。和一般服务器相比,I/O 操作要优越很多。

2. 卷管理器

卷管理器主要是进行磁盘和分区管理,包括磁盘监测、异常处理和逻辑卷的配置管理,一般应支持磁盘的热插拔、热替换等功能和 RAID 0、RAID 1、RAID 5 逻辑卷。

3. 文件系统

NAS 文件系统能支持多用户,具备日志文件系统功能,在系统崩溃或掉电重启后能迅速恢复文件系统的一致性和完整性。

4. 网络文件共享

网络文件共享一般支持 FTP、HTTP、NFS(UNIX)、CIFS(Windows 系统)、NCP(Novell 系统)、AFP(Apple 产品)等文件传输和共享协议。

5. Web 管理

Web 管理主要是包括远程监控和管理 NAS 设备的系统参数,例如网络配置、用户与组管理、卷以及文件共享权限等。

NAS 并不像普通 PC 系统一样拥有标准化的软件组件可供选择,NAS 厂商必须根据硬件自行设计软件系统。尽管设计工作是在已有资源的基础上进行的,但是难度也相当大。NAS 厂商必须对 Linux 内核进行裁剪或修改,使之可以在自身硬件上运行,同时必须自行编写设备的驱动程序和应用软件,工作量不亚于独立开发一套嵌入式操作系统。

为了提高系统性能并进行不间断的用户访问,NAS 常采用专业化的操作系统用于网络文件的访问,这些操作系统既支持标准的文件访问,也支持相应的网络协议。NAS 只保留了通用操作系统中用于数据共享的文件系统和网络连接协议,使 CPU、内存和 I/O 总线等系统资源完全用于信息资源的存储、管理和共享,使 NAS 的数据吞吐量比通用服务器大幅提高。

5.2.3 NAS 的实现架构

NAS 的实现架构分为集成式 NAS 架构和网关式 NAS 架构。集成式 NAS 的控制器和存储设备是使用 DAS 连接方式,而网关式 NAS 的控制器和存储设备是使用 SAN 连接方式。

1. 集成式 NAS 架构

集成式 NAS 架构为“NAS 头+SAN”。控制器部分俗称 NAS 头,是指将传统 NAS 机头的存储服务管理功能独立出来成为专门的存储服务器,通过基于 IP 或 FC 的接口和后端的存储区域网(SAN)直接相连,将访问频繁的数据存入 FC 磁盘,不太活跃的数据存入 SATA 磁盘,在一定程度上解决了成本和性能的矛盾,如图 5-4 所示。注意,后端的存储网络是非共享的(不需要经过交换机),可以将后端存储网络当成本地硬盘,将 IP 和 FC 接口当成总线接口,这也是集成式的 NAS 架构和网关式的 NAS 架构的主要差异。

2. 网关式 NAS 架构

网关式 NAS 架构为“NAS 网关+SAN”,是指将传统 NAS 机头的存储服务管理功能独立出来成为

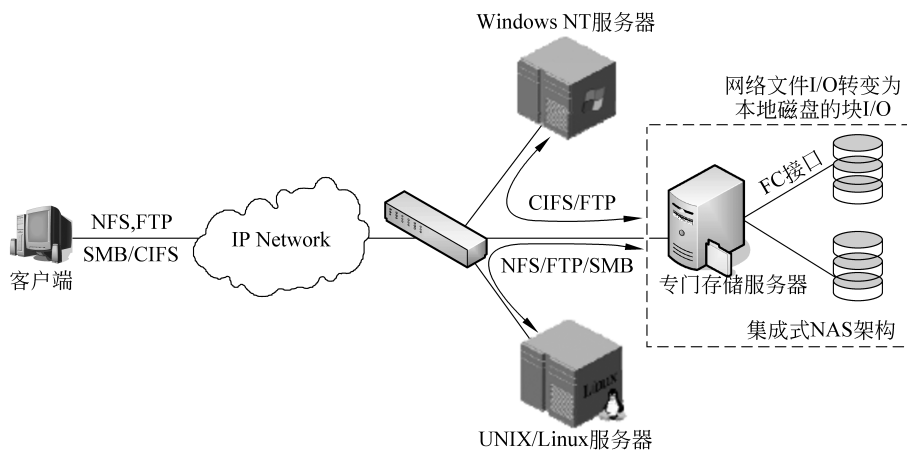


图 5-4 集成式 NAS 架构

专门的存储服务器,前端连接应用服务器设备,后端通过存储交换机(基于 IP 或 FC)和共享的 SAN 相连,初步实现 NAS 和 SAN 的融合,如图 5-5 所示。这也正迎合了目前各个主流存储厂商号称“统一存储”的提法。

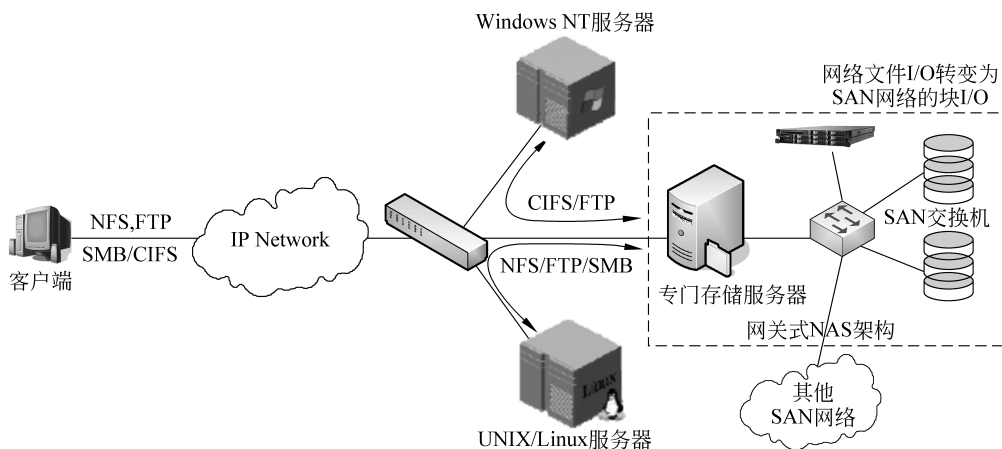


图 5-5 网关式 NAS 架构

网关式 NAS 是由一个独立的控制器即 NAS 头和若干存储阵列组成。NAS 头具有网关的功能,它可以访问 SAN 环境中的存储资源。这种方案的最大优势是,客户机可以通过 IP 网络访问 SAN 的存储资源,NAS 可以通过 NFS 和 CIFS 协议处理来自客户端的文件级 I/O 请求。当网关收到客户端发来的文件级 I/O 请求,便将该请求转化为块级 I/O 请求发送到 SAN 环境中的存储阵列,存储阵列处理该请求后又发给网关,网关将块级 I/O 转换成文件级 I/O,再发给客户端,对于客户端来说,整个过程是透明的。网关式 NAS 可以充分利用 SAN 存储资源,可扩展性很强,存储阵列和 NAS 头可以根据需要独立地扩展升级。

当存储资源出现紧缺时,可对连接 NAS 头的后端进行扩容,增添存储设备。当 NAS 的处理能力需加强时,NAS 处理器可以独立升级。总之,两者之间是相对独立。

无论是集成式 NAS 架构还是网关式 NAS 架构,都需要将传统 NAS 机头的存储服务管理功能独立出来成为专门的存储服务器,而 NAS 系统即为独立存储服务器的管理系统。

作为独立存储服务器的管理系统,NAS 系统承载着所有的存储服务管理功能。存储服务的管理主要涉及用户和用户组管理、底层逻辑卷资源管理、文件系统管理、共享文件夹管理、用户对资源的配额管理、用户对共享文件夹的权限管理、用户的接入认证管理、系统日志管理,以及系统状态监控等。

由于 NAS 系统一般是使用特别处理的剪裁后的操作系统,因此 NAS 的系统一般都没有其他的过于多余的功能,紧凑、高效,并且也将不必要的外部设备去掉,做到设备最大的精简,并对于存储文件、网络传输部分有所增强,如去掉显示设备、取消打印口、采用万兆以太网卡等。

5.3 NAS 协议

NAS 使用了传统的以太网协议,由于都是基于操作系统的文件共享协议,所以其性能特点适合进行小文件数量级的共享存取。提供各种应用领域的异种文件共享和文件服务功能,包括内容传送和分发、统一的存储管理、技术计算以及服务。它允许企业在不使服务器停机的前提下做扩容动作,甚至可以不停机进行物理迁移。

NAS 协议是用于实现网络文件系统功能的协议。网络文件系统的逻辑不是在本机运行,而是在网络上的其他远程主机结点上运行,本地端通过外部网络将读写文件的信息传递给运行在远端的文件系统。

5.3.1 NFS 协议

NFS(network file system,网络文件系统)是当前主流的异构平台共享文件系统之一,主要应用于 UNIX、Linux 环境,是一种实现 UNIX 系统之间磁盘文件共享的方法。NFS 协议是一种支持客户端的应用程序通过网络存取服务器磁盘中数据的一种文件系统协议,最早是由 Sun 和 IBM 公司共同开发,现在已支持在不同类型的系统之间通过网络进行文件共享,被广泛应用于 FreeBSD、SCO、Solaris 等异构操作系统通过网络彼此共享目录和文件。

通过使用 NFS,用户和程序可以像访问本地文件一样远程访问其他系统上的文件,使每个计算机结点都能像使用本地资源一样方便地使用网络上的资源。换言之,NFS 可实现操作系统、网络架构和传输协议等不同的计算机在网络环境下进行文件远程访问和共享。

5.3.2 SMB/CIFS 协议

CIFS(common Internet file system,通用互联网文件系统)是当前主流异构平台共享文件系统之一,主要用于在 Windows 环境中连接 Windows 客户机和服务器。CIFS 是微软公司整合了 SMB(server message block,服务器信息块)协议后的成果,可使程序远程访问 Internet 上计算机的文件并申请此计算机的服务。CIFS 可与支持 SMB 的服务器进行通信,实现共享。因此,为了让使用 Windows 操作系统的计算机可以和使用 UNIX、Linux 操作系统的计算机相互进行文件访问,CIFS 协议是必要的。

CIFS 采用客户-服务器模型,是在 Windows 系统中部署 NAS 设备时最常用的两种协议之一。其工作原理是让 CIFS 协议运行于 TCP/IP 之上,客户端请求远程服务器上的服务器程序为它提供服务,服务器收到请求并返回响应。

5.3.3 文件传输协议

文件传输协议(FTP)用于在 Internet 上控制文件的双向传输。FTP 的主要作用是让用户连接一个远程计算机并查看该机上的文件,然后将其从远程计算机下载到本地计算机,或把本地计算机的文件上

传到远程计算机。

NAS 系统最核心的功能是对外提供文件共享服务,因此几乎所有与 HTTP(超文本传送协议)等文件传输相关的协议均可作为 NAS 协议应用于 NAS 系统中。

异构平台之间的文件共享服务需要网络文件系统的支持,网络文件系统共享协议(如 NFS、SMB、CIFS)是 NAS 协议的根本和基础,FTP、HTTP 等其他文件传输相关的协议只是应用的补充。

5.4 NAS 文件系统

虽然 NAS 设备常被认为是一种存储架构,但是 NAS 设备最核心的功能是文件管理服务。至于通过 NFS/CIFS 共享文件,完全属于高层协议通信,根本就不在数据 I/O 路径上,所以数据的传输不可能以块来组织。正是由于这种功能上的重叠,在 SAN 出现以后,NAS 头(或 NAS 网关)设备逐渐发展起来,NAS over SAN 的方案越来越多,NAS 回归了其文件服务的本质。

由此可知,NAS 与一般的应用主机在网络层次上的位置是相同的,为了在磁盘中存储数据,就必须建立文件系统。有的 NAS 设备采用专有文件系统,而有的 NAS 设备则直接借用其操作系统支持的文件系统。由于不同的操作系统平台之间文件系统不兼容,所以 NAS 设备和客户端之间就采用通用的 NFS 或 CIFS 来共享文件。

NAS 服务器可以看成是一个专门的文件管理器,主要由专为提供文件服务而优化的操作系统、相关系统软件和相关硬件组成。图 5-6 所示为 NAS 数据访问层次,前端应用服务器(UNIX、Linux 或

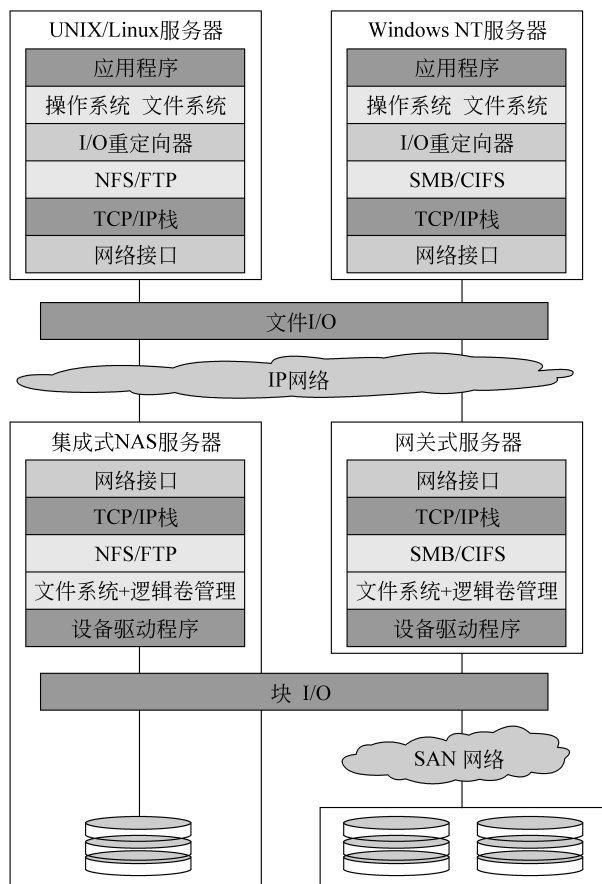


图 5-6 NAS 数据访问层次

Windows NT 服务器)利用网络文件系统(NFS、SMB 或 CIFS)映射 NAS 服务器端(即存储端)的本地文件系统(Ext3、Ext4、ReiserFS、XFS、JFS 等)到应用服务器端自己的操作系统中,这样应用服务器端就可以像访问自己的本地磁盘空间一样访问远程 NAS 服务器端的存储空间。

NAS 通过 NFS 和 CIFS 协议管理远程文件系统上的 I/O 请求处理,NAS 的 I/O 过程与块级的 I/O 过程不同。首先,请求者把一个 I/O 请求封装成 TCP/IP 报文,在网络协议栈转发,直到 NAS 设备从网络上接受这个 I/O 请求。其次,接到这个 I/O 请求后,NAS 设备将它解包转换成一种块级请求,即一种对应的物理存储请求,然后对物理资源存储池进行相对应的操作。最后,数据从物理存储池返回时,NAS 设备将块级的 I/O 封装成相应的文件协议然后把文件协议与相应的操作封装在 TCP/IP 中,最后通过网络转发出去。

5.5 NAS 资源

文件系统是操作系统用于在存储介质上组织和管理文件的方法,是负责管理和存储文件信息的软件系统。文件系统的载体为存储介质,在 NAS 系统中即为 NAS 资源,NAS 系统支持的 NAS 资源主要有磁盘及分区、磁盘阵列,以及逻辑卷。

5.5.1 磁盘及分区

传统 NAS 机头的存储池一般为直接用 IDE、SATA 或 SCSI 总线连接的本地磁盘。本地磁盘首先需要经过分区,然后每个分区才能被格式化为指定的文件系统为 NAS 系统所用。一个磁盘,既可以整体划分为一个分区,也可以划分为多个分区,分区的划分需要考虑“3P+1E”的原则(即 3 个主分区+1 个扩展分区,此处的 P 代表 primary,E 代表 extended)。

由于目前单个磁盘的容量已经突破了 1TB,所以简易的 NAS 系统可以只是独立的存储服务器,直接利用本地磁盘作为 NAS 资源,而不用后端的 SAN 存储池构建的磁盘阵列。

5.5.2 磁盘阵列

在网关式 NAS 架构中,NAS 服务器的后端通过存储交换机(基于 IP 或 FC)和共享的 SAN 相连,SAN 网络的存储池提供给 NAS 主机的,实际就是一些未建立文件系统的“虚拟磁盘”。这里的“虚拟磁盘”为根据不同的 RAID 策略,采用多块物理磁盘组成的磁盘阵列。

RAID 技术可将多个单独的磁盘以不同的组合方式形成一个逻辑硬盘,从而提高磁盘读取的性能和数据的安全性。不同的组合方式用 RAID 级别来标识。

RAID 技术经过不断的发展,现已拥有 RAID 0~RAID 5 共 6 种 RAID 级别标准。另外,其他还有 6、7、10(RAID 1 与 RAID 0 的组合)、01(RAID 0 与 RAID 1 的组合)、30(RAID 3 与 RAID 0 的组合)、50(RAID 0 与 RAID 5 的组合)等级别,不同 RAID 级别代表着不同的存储性能、数据安全性和存储成本。

5.5.3 逻辑卷

逻辑盘卷管理是 Linux 环境下对磁盘分区进行管理的一种机制,具有强大磁盘管理功能,它巧妙地吧传统的物理分区转换成逻辑分区,使管理人员从传统的分区管理中彻底解放出来。通过 LVM 系统,逻辑卷管理将若干个磁盘分区连接为一个整块的卷组,形成一个存储池。管理员在卷组上随意创建逻辑卷,可以方便实现文件系统跨越不同磁盘和分区。

无论是磁盘还是磁盘阵列对象,若直接对其划分物理分区,则将来要在这些分区上新增或减少磁盘

空间会非常麻烦;此外,物理分区的读写模式是线性模式,性能也不高。

逻辑卷的思想就是将多个物理磁盘分区,甚至磁盘阵列对象组合起来,形成一个独立的虚拟大磁盘,然后在虚拟磁盘空间上动态地划分需要的空间建立逻辑卷。

目前,虚拟磁盘技术和 LVM(逻辑卷管理)技术都支持逻辑卷,如图 5-7 和图 5-8 所示。不同的是,LVM 最大的用途是灵活地管理磁盘的容量,让逻辑分区可以随时放大或缩小,便于更好地利用剩余的磁盘空间;而虚拟磁盘技术由于以 RAID 为基础,更强调性能与备份。

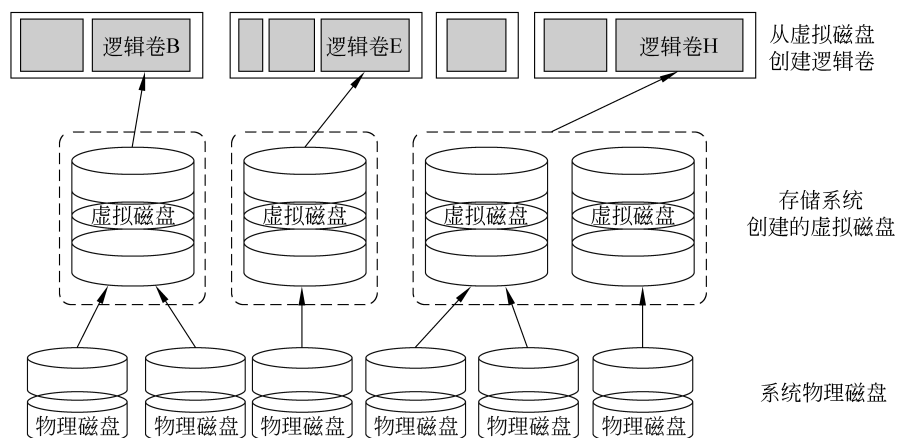


图 5-7 从虚拟磁盘创建逻辑卷

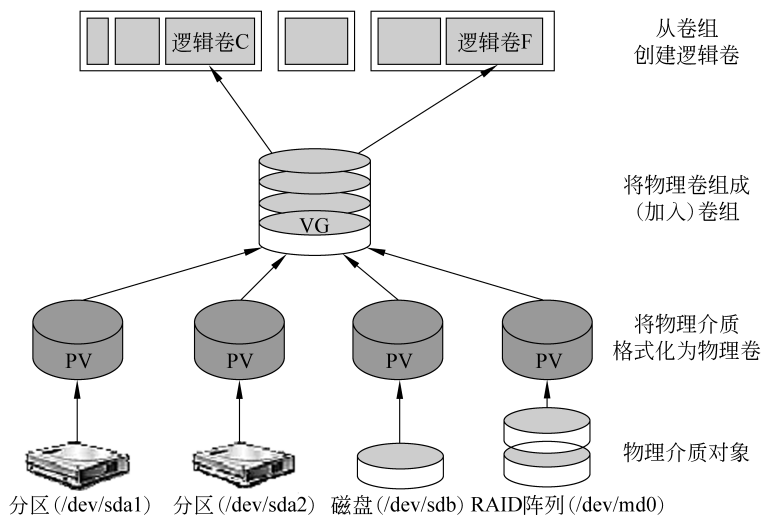


图 5-8 由 LVM 创建逻辑卷

5.6 集群 NAS 存储系统

目前的存储设备中,NAS 技术的使用存在一定的局限性。首先,NAS 存在性能缺陷,在 NAS 系统中,对 NAS 的请求和对应用服务器的请求将很有可能同时竞争有限的带宽,即使在万兆以太网中,这种冲突依然存在;其次,NAS 是基于文件访问的,它很难做到像 SAN 那样基于块设备的高速访问。

NAS 很难避免“信息孤岛”的出现,如果在一个企业中存在多个 NAS,很有可能在一个 NAS 上的

数据同时出现在另一个 NAS 上,并与此同时,各个 NAS 之间无法做到相互沟通,相互共享资源。

NAS 容易出现单点失效的局面,由于 NAS 没有像 SAN 那样在链路上的冗余,不能有效避免单点失效。NAS 集群技术可以一定程度解决这些问题,用多台 NAS 组成一个大容量、高可用性、高性能、高扩展性的存储系统。

集群(cluster)是由多个结点构成的一种松散耦合的计算结点集合,协同起来对外提供服务。集群 NAS 是指协同多个 NAS 结点(即通常所称的 NAS 机头)对外提供高性能、高可用和高负载均衡的 NAS(NFS、CIFS、FTP)服务。有别于传统的 SAN 和 NAS,集群 NAS 是一种横向扩展(scale out)存储架构,主要面向文件级别的存储,具有独立升级、线性扩展(容量增加不会带来性能瓶颈)的优势,已经得到了全球市场的广泛认可,成为了主流的存储技术之一。

集群 NAS 的整体架构由存储子系统、NAS 机头(集群)、客户端以及网络组成。存储子系统可以采用存储区域网(SAN)、直接附接存储(DAS)或者面向对象存储设备(OSD)的存储架构。根据所采用的后端存储子系统的不同,可以把集群 NAS 分为 SAN 共享存储架构、集群文件系统架构和 pNFS/NFS v4.1 架构 3 种技术架构。

5.6.1 SAN 共享存储架构

在 SAN 共享存储架构中,前端的 NAS 结点通过光纤连接至后端 SAN 共享存储设备,采用 SAN 并行文件系统管理数据并对外提供 POSIX 接口,其架构如图 5-9 所示。元数据则有专用元数据服务器(MDC)和分布到 SAN 客户端两种方式。要实现对 SAN 共享存储的并发访问,只需要在集群 NAS 上安装 SAN 文件系统客户端即可,用户客户端的数据访问通过 NFS 或 CIFS 协议进行。

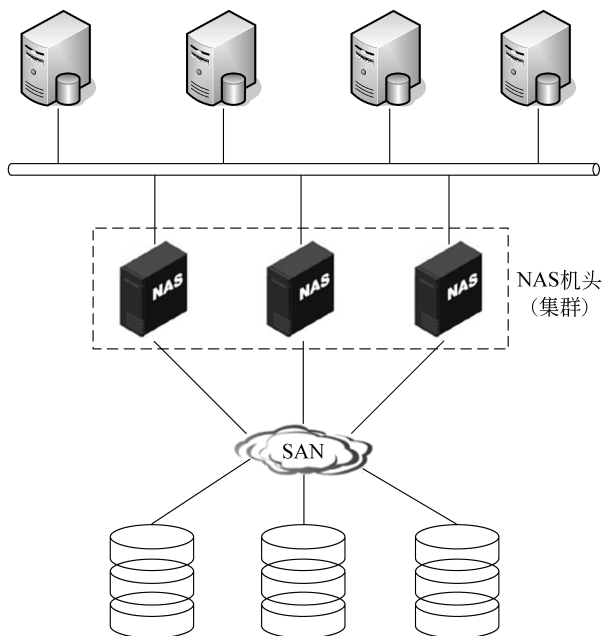


图 5-9 SAN 共享存储集群 NAS 架构

由于 SAN 共享存储架构是基于高性能的 SAN 存储,所以这种架构可以提供稳定的带宽和性能,并通过存储磁盘阵列和 NAS 结点的添加进行拓展。SAN 共享存储架构的缺点是,SAN 存储网络和并行文件系统的成本都比较高,SAN 架构本身还存在部署管理复杂、拓展规模有限的问题。