

数字图书馆建设之存储架构分析

刘冬邻^{1,2} 陈建国² 许宏玲³

(重庆大学计算机学院 重庆 400044)¹(四川外语学院 重庆 400031)²(西南政法大学 重庆 400031)³

摘要 本文介绍了现在主要的存储架构以及存储分类,分析了 DAS、NAS、SAN 存储架构的结构,支撑技术以及各自的特点,为数字图书馆的存储建设提供了架构选择的参考。

关键词 数字图书馆,存储架构,DAS,NAS,SAN,FC-SAN,IP-SAN

1 引言

随着信息技术和网络技术的发展以及各高校自身建设的需要,越来越多的高校把数字图书馆的建设作为学校数字化校园建设的一个非常重要的组成部分。而存储系统的建设又是整个数字图书馆建设的重要的基础设施建设,它的建设效果将直接影响到相关的应用建设,影响到整个数字图书馆系统的水平和性能。存储系统的建设首先是存储架构的选择。当前的各个厂家和系统集成商都能够为我们提供众多的存储系统方案和存储系统架构供我们选择,这为数字图书馆的建设提供了发展的机会,也给系统的建设者们带来了困惑。我们在建设数字图书馆的存储系统时如何来选择合适的存储架构? 本文将对现在主要的存储架构进行分析。

2 当前主要的存储架构及其特点

目前磁盘存储的存储分类(如图 1)根据服务器类型分为:封闭系统的存储和开放系统的存储,封闭系统主要指大型机等服务器,开放系统指基于包括 Windows、UNIX、Solaris 等操作系统的服务器;开放系统的存储分为:内置存储和外挂存储;开放系统的外挂存储根据连接的方式分为:直连式附加存储(Direct-Attached Storage,简称 DAS)和网络化附加存储(Fabric-Attached Storage,简称 FAS);开放系统的网络化存储根据传输协议又分为:网络接入存储(Network-Attached Storage,简称 NAS)和存储区域网络(Storage Area Network,简称 SAN)。目前绝大部分采用的是开放系统。

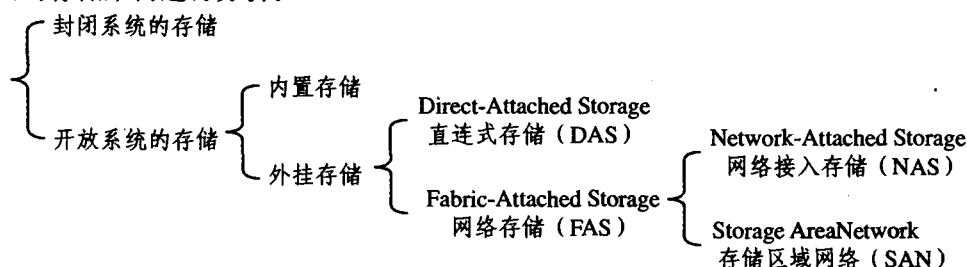


图 1 存储分类

2.1 DAS(Direct Attached Storage)存储架构的特点

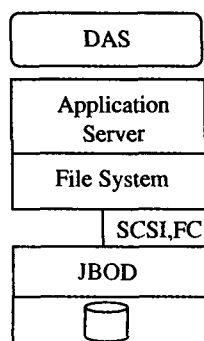


图 2 DAS存储架构

DAS即“直接连接存储”,指将外置存储设备通过 SCSI 接口或光纤通道,直接连接到一台计算机上。采用直接外挂存储方案的服务器结构如同 PC 机架构,外部数据存储设备采用 SCSI 技术,或者 FC(Fibre Channel)技术,直接挂接在内部总线上的方式,存储设备一般与服务器是一对一。数据存储是整个服务器结构的一部分。

直连式存储(DAS)已经有近四十年的使用历史,随着用户数据的不断增长,尤其是数百 GB 以上时,其在备份、恢复、扩展、灾备等方面的问题变得日益困扰系统管理员。用户的数据访问必须通过服务器,经过 I/O 总线访问相应的存储设备,服务器起

到资源存储转发的作用。直连式存储与服务器主机之间的连接通道通常采用 SCSI 连接,带宽为 40MB/s、80MB/s、160MB/s 等,随着服务器 CPU 的处理能力越来越强,存储硬盘空间越来越大,阵列的硬盘数量越来越多,SCSI 通道将会成为 IO 瓶颈;服务器主机 SCSI ID 资源有限,能够建立的 SCSI 通道连接有限。DAS 的系统安装、RAID 调试比较烦琐,无独立的存储操作系统,需相应服务器或客户端支持,依赖服务器对其进行管理和维护,存储设备仅起到数据的存储作用,如果服务器出现问题,容易造成管理和维护瘫痪。不过现在的硬件阵列设备一般采用 RAID 5 技术,存储数据本身的安全性还是比较可靠的,但在增加硬盘时,需要重新做 RAID,这时须宕机,会影响服务。DAS 实行异地备份,备份过程麻烦。依靠双服务器和相关软件实现双机容错功能,直连式存储依赖服务器主机操作系统进行数据的 IO 读写和存储维护管理,数据备份和恢复要求占用服务器主机资源(包括 CPU、系统 IO 等),数据流需要回流主机再到服务器连接的存储设备,数据备份通常占用服务器主机资源 20%~30%,直连式存储的数据量越大,备份和恢复的时间就越长,对服务器硬件的依赖性和影响就越大。

2.2 NAS (Network Attached Storage) 存储架构的特点

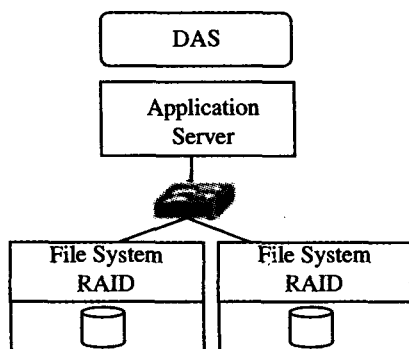


图3 NAS 存储架构

NAS,网络附加存储也叫网络连接存储,它是一种集 CPU、内存、操作系统和磁盘系统于一体的性能较高、跨平台的数据存储技术,采用网络(TCP/IP、ATM、FDDI)技术,通过网络交换机连接存储系统和服务器主机,建立专用于数据存储的存储私网。它是一种向用户提供文件级服务的专用数据存储设备,直接连到网络上,不再挂接服务器后端,避免给服务器增加 I/O 负载。NAS 产品将所有的软件都全部固化在自身设备内,支持目前流行的各种协议(包括 TCP/IP、IPX/SPX 等)。

NAS 架构能用来实现在不同操作系统平台下的文件共享应用以及网络磁盘,用户空间限制等功能,与传统的服务器或 DAS 存储设备相比,NAS 设

备的安装、调试、使用和管理非常简单,采用 NAS 可以节省一定的设备管理与维护费用。NAS 设备提供 RJ-45 接口和单独的 IP 地址,可以将其直接挂接在主干网的交换机上,通过简单的设置就可以在网络中即插即用使用其提供的存储服务,而且进行网络数据在线扩容时也无需停顿,从而保证数据流畅存储。

NAS 目前采用的协议是 NFS 和 CIFS。NFS (Network File System) 协议是由 Sun 公司开发的基于 Unix 环境下的网络文件系统。CIFS (Common Internet File System) 则是由微软开发,用于连接 Windows 客户机和服务器的通用互联网文件系统。

随着 IP 网络技术的发展,网络接入存储(NAS)技术也是日新月异。20 世纪 80 年代末到 90 年代初的 10Mbps 带宽,网络接入存储作为文件服务器存储,性能受带宽影响;后来快速以太网(100Mbps)、VLAN 虚网、Trunk (Ethernet Channel) 以太网通道的出现,使网络接入存储的读写性能得到改善;1998 年千兆以太网(1000Mbps)的出现,为网络接入存储(NAS)带来质的变化和广泛认可。由于网络接入存储采用 TCP/IP 网络进行数据交换,TCP/IP 是 IT 业界的标准协议,不同厂商的产品(服务器、交换机、NAS 存储)只要满足协议标准就能够实现互连互通;并且随着 2002 年万兆以太网(10000Mbps)的出现,存储网络带宽将大大提高 NAS 存储的性能。

NAS 与传统的以太网相连,使用的是 TCP/IP 协议,当需进行文件共享和访问时则利用 NFS 和 CIFS 以沟通 Windows NT 或 Unix 系统。NAS 的这种结构奠定了它自身的优点,那就是配置灵活、管理方便,另外成本较低,不需要配置额外的网络配件。但因为 NAS 是直接和网络相连的,无疑在数据传输时会增加网络负担,一旦发生网络拥塞,其性能会大幅下降。所以 NAS 自身要受到网络传输能力的制约。

2.3 SAN (Storage Area Networks) 存储架构的特点

SAN,即“存储区域网络”,是在存储设备和服务器之间建立直接的高速专用网络连接,通过这种连接实现只受光纤线路长度限制的集中式存储。SAN 可以被看作是存储总线概念的一个扩展,它使用 LAN/WAN 中类似的单元,该存储结构是一个分布式、专用网络,用于高效地传输大量数据。它由存储池系统、服务器群、FC 光纤通道、文件管理系统四个主要组成部分。存储池系统由 FC 硬盘存储器组成。大型数字图书馆数据要求硬盘具有非常大的带宽。FC 光纤通道由其核心部件光纤交换

机、光纤网卡和光纤或铜缆组成,本质上是 SC2SI 标准的串行化。目前常用的 FC 不仅可采用光纤,也可采用铜线传输,只有远距离传输时,才必须使用光纤。光纤通道是一种高速(1-4Gb/s)串行接口(现已发展到 10Gb/s),它支持多种拓扑和协议。SAN 采用的光纤通道拓扑是一种可交换的网络结构,它可以在网络中任意成对节点之间通过光纤交换机做多个点对点连接。在 SAN 中,服务器和硬盘器件都被看作是一个节点,在任一时间每对服务器和硬盘都可以共享光纤通道的全部带宽。服务器群由多台服务器组成,可以根据实际需求分散安置。现在服务器多带有 FC 接口,可方便地联接在 SAN 网络中。由于采用了光纤连接,服务器可相距很远。每个服务器可根据系统情况,装载不同数量的数字资源。文件管理系统管理整个网络(包括文件管理服务器和以太网),以太网主要用于控制服务器、数据服务器、客户三者之间的通信。

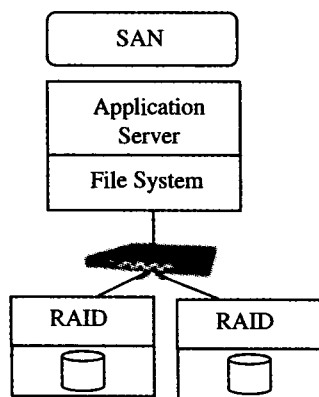


图 4 SAN 存储架构

SAN 的支撑技术是光纤通道——Fibre Channel(FC)技术,FC 是 ANSI 为网络和通道 I/O 接口建立的一个标准集成。支持 HIP PI、IPI、SCSI、IP、ATM 等多种高级协议,它的最大特性是将网络和设备的通讯协议与传输物理介质隔离开。这样多种协议可在同一个物理连接上同时传送,高性能存储体和宽带网络使用单 I/O 接口,使得系统的成本和复杂程度大大降低。光纤通道支持多种拓扑结构,主要有:点到点(Links)、仲裁环(FC-AL)、交换式网络结构(FC-XS)。

光纤通道(Fibre Channel)SCSI 技术是 SAN 技术的物理基础。Fibre Channel 采用高频(1GHz, 2GHz, 4GHz)串行位(Bit)传送,单环速度可达 100~200Mbyte/s(相当于 Gigabit),双环共用可达到 200~400Mbyte/s 甚至更高的速率。每个环可挂接 126 个 SCSI 设备,不加中继时最远距离可达 10Km,而且有很大的继续发展空间。传统的 SCSI 总线电缆因受制于电子技术和电气物理特性的限制,在速度(20~160MB/s),容量(每条总线 8~16

个 SCSI 设备),距离(1.5~25 米)等方面都已近极限。

Fibre Channel 采用 FC-AL 仲裁环机制,使用 Token(令牌)的方式进行仲裁,其效率远较传统 Ethernet 的 CSMA/CD 为高;另外,SAN 的网络协议为 SCSI-3,在数据流的包/帧结构上,其效率远较 TCP/IP 为高。SAN 不仅保留了传统的 RAID, HA, Cluster 等安全措施,而且提供了双环冗余,远程备份等新的安全手段。

3 SAN (Storage Area Networks)存储架构的 IP 实现

现在很多高校在建设数字图书馆时,在计划自己的 SAN 中普遍使用光纤通道方式即 FC-SAN。能够以合理价位替代光纤通道的存储网络产品目前尚不太多。使用千兆以太网和基于 IP 的存储技术将成为必然之路。使用通用的网络产品来实施存储网络可以少花钱,网络管理人员也不需要额外培训。除光纤通道外,已经成熟起来的 IP 存储将会是不错的选择。

IP 存储就是使用 IP,而不是光纤通道,把服务器与存储设备连接起来的技术,除了已获通过的 iSCSI 标准,还有 FCIP、iFCP 等正在制定的标准。而 iSCSI 发展最快,已经成了 IP 存储的代表。2003 年 2 月 11 日,IETF 通过了 iSCSI 标准。这将吸引更多的人参与到相关产品的开发中,也会推动更多的用户采用 iSCSI 解决方案。

iSCSI 实现了 SCSI 和 TCP/IP 协议的连接,该技术允许用户通过 TCP/IP 网络来构建存储区域网(SAN)。iSCSI 可提供必要的映射,通过 IP 传输 SCSI 指令,就像今天的光纤通道可以传输 SCSI 指令一样。iSCSI 是为主机到存储设备的端到端连接而设计的,类似于光纤通道的 SAN 架构,iSCSI 技术可使主机到兼容的存储设备之间通过 IP 交换机进行通讯。而驱动器仍可使用真正的 SCSI 驱动器,因为 iSCSI 并不等同于今天的硬盘连接技术。

采用 IP 存储方案,只要联入局域网的计算机都可以使用存储系统里的存储空间,使得整个存储系统具有高扩展性、高性价比、低 TCO,采用 iSCSI 标准,只要计算机有网卡,就可以使用存储系统中的存储空间,以实现低成本的 IP-SAN。采用互为热备的存储设备,光纤磁盘阵列采用双控制器,而且两台磁盘阵列互为备份;两台存储交换机进行群集,提供双机热备负载均衡,从而保证整个系统的高可靠性。部分重要应用的服务器进行双机工作,从而保证关键业务不被中断,提供系统的可靠性和稳定性。

结论 DAS 这种架构,能够解决单台服务器的

(下转第 156 页)

得到 a, b, c 的关系应为 $a=b, c=0$, 但是由于实际问题中总是存在误差, 因此得到的情况是 $a>b>c$, 于是令 $\tilde{D}=\text{diag}((a+b)/2, (a+b)/2, 0)$, 重新得到新的本质矩阵 $\tilde{E}=U\tilde{D}V^T$, 对 E 再进行奇异值分解, $\tilde{E}=\tilde{U}\tilde{K}\tilde{V}^T$, 这样就可以得到下面照相机运动参数 $(R|T)$ 的候选值: $(\tilde{U}\tilde{K}\tilde{V}^T|\tilde{U}(0,0,1)^T)$ 。

(3) 计算空间点三维坐标

在获得运动参数 $(R|T)$ 后, 给定一对匹配对, 可以直接利用线性最小二乘法求出相应的空间三维点坐标。

于是根据最小二乘法得

$$X=(Z^T Z)^{-1} Z^T z \quad (12)$$

对每一个匹配对采用如上方法可以获得所有的空间点的三维坐标。

6 三维点阵配准拼接

假定来自一个物体上的两个点集(或两个网面)分别表示为 A 和 B , 其中 A 和 B 在物体表面上有足够的重叠以便实现可靠对准。点集 B 中的一点 q 和点集 A 的距离是

$$d(q, A)=\min_{p \in A} \|q-p\| \quad (13)$$

点集 B 中的一个子集 $Q=\{q_i\}$ 与点集 A 中的一个子集 $P=\{p_i\}$ 构成最近距离点对, 形成共轭对集 $\{(p_1, q_1), (p_2, q_2), \dots, (p_n, q_n)\}$, 这个共轭对集可以用来求解绝对定位问题。如果这些点集不是真正的共轭对, 也可以获得两个网面之间变换的逼近解法。两个网面之间变换以后, 重新求解最近共轭点集, 使得该共轭对集更接近真正的共轭对集, 这样的过程可以重复进行, 直到最近点之间的距离均方差低于某一阈值。

结论 本文首先介绍立体视觉三维重建的基本原理, 然后介绍了视觉系统中的一个重要约束条

件——极线约束。由于欧式几何意义下的重建要求的条件过多(如需要严格的照相机定标等), 而有些信息在某些场合下是不必要的, 因此就提出了射影几何意义下的重建, 可以做到只在解决基本矩阵 F 的照相机弱标定条件下, 得到识别物体所需要的射影不变量, 从而既简化了计算和对系统的要求, 又满足相当一部分视觉任务的需要。本文使用的重建算法收敛的条件是为对应图像中点对的相对误差范数小于一个临界值。但是需要指出的是, 由于点云数据中的噪声影响, 若干误差比较大的点对于配准拼接算法精度有很大的影响, 鲁棒性方面有待提高。配准拼接算法和目前普遍采用的方法及 Reich 提出的集成测量方法相比, 既不需要精密机械定位和光学跟踪定位, 其精度也不完全取决于标志点视觉测量。

参 考 文 献

- 1 Hartley R, Gupat R, Chang T. Stereo from uncalibrated cameras [J]. In: Proc. IEEE Conf. on Computer vision and pattern Recognition. 1992. 761~764
- 2 Ganci G, Brown J. Development in non-contact measurement using video grammetry [J]. In: Boeing Large Scale Metrology Conf. Long Beach, 2000
- 3 Moravec, Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover by H. Moravec [J]. Stanford AI memo AIM-340, 1980
- 4 Harris C, Stephens M. A combined corner and edge detector [J]. In: the Fourth Alvey Vision Conference, 1988
- 5 Zhang Z Y. Flexible camera calibration by viewing a plane from unknown orientations [J]. In: International Conference on Computer Vision (ICCV99), Greece, 1999. 666~673
- 6 孟晓桥. 摄像机自标定和三维重建中的若干问题研究[D]. [中国科学院研究生院硕士论文]. 2000
- 7 马颂德, 张正友. 计算机视觉[M]. 科学出版社, 1998

(上接第 118 页)

存储空间扩展、高性能传输需求, 并且单台外置存储系统的容量, 已经发展到了 2TB 以上, 随着大容量硬盘的推出, 单台外置存储系统容量还会上升。此外, DAS 还可以构成基于磁盘阵列的双机高可用系统, 满足数据存储对高可用的要求。从趋势上看, DAS 仍然会作为一种存储模式, 继续得到应用。

SAN 对于高容量块状级数据传输具有明显的优势, 而 NAS 则更加适合文件级别上的数据处理。尽管二者存在根本特性上的差异, 但 SAN 和 NAS 实际上也是能够相互补充的。SAN 擅长块数据传输、极易扩展且管理设备有效。用户可以使用 SAN 运行关键应用, 比如数据库、备份等, 以进行数据的集中存取与管理; 而 NAS 支持若干客户端之间文

件共享, 所以我们可以使用 NAS 作为日常办公中需要经常交换小文件的地方, 比如网页、公文等。SAN 和 NAS 在实际情况中是可以并存在一个系统中的, 从而构建一个功能强大, 结构灵活的存储系统。

参 考 文 献

- 1 袁文莉. 数字图书馆存储体系架构探析. 图书馆论坛, 2004(8)
- 2 SAN 与 NAS 和谐统一. 网络世界, 2005. 10. 13
- 3 存储在线. 深入了解网络存储, SAN、NAS 技术的比较. PConline, 2003. 10. 30
- 4 杨子伍. 探讨数字化图书馆的存储策略. 现代图书情报技术, 2003(3)
- 5 高先锋, 张洪滔. 图书馆海量存储系统架构与接口的选择. 现代图书情报技术, 2003(4)