



信息技术应用系列

实用网络存储技术

顾锦旗 胡苏太 朱平 编著



上海交通大学出版社

信息技术应用系列

实用网络存储技术

顾锦旗 胡苏太 朱平 编著

上海交通大学出版社



内 容 提 要

在信息化时代, 存储技术正在发生着革命性变化。这种变化主要体现在“信息爆炸”时代给信息存储和应用带来的挑战。存储区域网(SAN)是在这种形势下提出的一种新的存储技术, 它不仅能为网络上的应用系统提供丰富、快速、简便的存储资源, 而且还能对网上的存储资源实施集中统一的管理, 从而成为当今理想的存储管理和应用模式。本书将围绕 SAN 技术的联网原理、拓扑结构、SAN 管理与应用, 着重介绍光纤通道等 SAN 的主要实现技术, 并将分析存储技术的发展。

本书共分 12 章, 内容包括: 存储技术、高性能接口技术、存储器与联网概念、光纤通道的内部结构、SAN 拓扑、SAN 中问题的隔离、SAN 的管理、光纤通道产品、应用研究、新的存储联网技术和世界主要厂商的发展战略等。本书图文并茂, 突出实用性和技术性, 既可作为计算机网络专业人员的参考书, 也可作为高等院校计算机专业的教材。

图书在版编目(CIP)数据

实用网络存储技术 / 顾锦旗编著. — 上海: 上海交通大学出版社, 2002(信息技术应用系列)
ISBN7-313-02961-6

I. 实... II. 顾... III. 计算机网络-信息存储
IV. TP393.0

中国版本图书馆 CIP 数据核字(2002)第 001811 号

实用网络存储技术

顾锦旗等 编著

上海交大出版社出版发行

(上海市番禺路 877 号 邮政编码 200030)

电话: 64071208 出版人: 张天蔚

上海锦佳装璜印刷发展公司印刷 全国新华书店经销

开本: 787mm×1092mm 1/16 印张: 11.5 字数: 277 千字

2002 年 2 月第 1 版 2002 年 2 月第 1 次印刷

印数: 1~3050

ISBN7-313-02961-6/TP·497 定价: 16.00 元

版权所有 侵权必究

前 言

在信息化时代，存储技术正在发生着革命性变化。这种变化主要体现在“信息爆炸”时代给信息存储和应用带来的挑战。

存储区域网（SAN）代表着从传统服务器到磁盘数据模型的主要变化。在大多数企业网络和信息网络中，文件服务器位于数据存取的中心。数据本身以各种形式存在，如电子邮件、字处理文档、电子表格、工程图、数字视频等等，而且通常都常驻在磁盘或光盘上。受最终用户请求的驱动，服务器永远是读磁盘上的数据文件和把数据文件写到磁盘上，并用适当的网络协议给数据打包，以便传送给用户。

传统数据存取模型要求每个数据请求都要通过文件服务器。文件服务器实际上拥有其附属磁盘上的全部数据。在网络一侧，服务器是网络生成的数据请求的集中点；在存储器一侧，服务器是为服务这些请求所需要的所有集中点。

这个老模型是建立在小型计算机系统接口（SCSI）并行总线结构上的，它用固定的专用连接把服务器连到存储器阵列上。特定磁盘上的数据在正常情况下只能由单一的服务器经 SCSI 总线进行存取。可供该服务器使用的数据量又受总线支持的磁盘数量的限制，总线的数量由服务器支持。

如果服务器或其同磁盘的任何 SCSI 连接发生故障，对数据的存取也随之丢失。对负有重要使命的网络来说，这种突然损失的潜在问题是不能接受的。对企业来说，存取数据受到的限制如同拨号音，存取的损失最直接地就是业务的损失。这种以服务器为中心的模型不适合于企业网络的高可用性和高容量要求这一事实，是改变服务器/存储器关系的主要动力。

存储区域网处在这种变化的最前沿。SAN 把联网的灵活性引入服务器和存储器，并通过去掉服务器和磁盘之间的专用连接，使以数据为中心的整体能够扩展。SAN 的速度、容量和网络灵活性主要基于光纤通道结构。光纤通道提供可扩展带宽、冗余数据路径、长距离（达 10km）和大规模设备群。

像出现的任何基础结构一样，存储区域网也具有很长的孕育期。在孕育期，该技术仅以建议、标准草稿和少数早期市售产品的形式存在。例如，光纤通道最初是在 20 世纪 80 年代末期提出，而自 1995 年以来作为推荐的 ANSI（美国国家标准局）标准而存在。第一批光纤通道产品问世于 1996 年。现在，有几十家公司在生产特别适合 SAN 市场的适配器卡、无线电收发机、集线器和开关。到 1999 年春季为止，文件服务器、磁盘存储器和磁带机的所有大销售商都在发售 SAN 使能（SAN-enabled）的产品。到 2002 年，SAN 市场可望超过 20 亿美元。

在 SAN 出现之前，服务器和存储器之间的连接避开了联网对数据通信其余部分展开的论战。路由器、开关、路由选择协议、任意连接性、网络拓扑结构等等在传统服务器/存储器模型中是一些不同性质的概念。只要请求数据的用户网络比服务器检索数据和把数据存入磁盘的能力慢，就无需重新设计服务器到存储器的链路。

但是，当用户网络越来越多，越来越快，并把更多的要求加到服务器资源上的时候，

就有问题了。服务器技术的竞争性能现估计为数万个 I/O 事务、万亿个字节数据量和每秒数百兆字节的服务器/存储器数据速率。把大量数据在合适的时间内备份到磁带上或其他媒体上已越来越困难。多媒体应用给存储器和带宽施加了一些新的要求。加速市场接受 SAN 概念和产品的速度显示出这些问题在企业网络中普遍存在，而光纤通道 SAN 正在提供一些可行的解决方法。

用 SAN 技术解决复杂存储器问题是个重要任务。存储器网络的设计要求仔细分析存储器的应用和 SAN 部件按要求执行的能力。在光纤通道标准形态下存在的特性不可以在实际产品形态下存在。这些标准本身可以定义某些与现实世界没有什么直接关系的抽象能力。遗憾的是，根据个别销售商的销售策略，一些 SAN 产品生产厂家为谋取浅薄的市场优势，可能过分夸大他们产品的能力。

在设计和实现 SAN 时，第一步是了解什么是现实的和什么是不现实的。例如，光纤通道往往被描述成 200Mbps 全双工传输。实际上，没有一种全双工应用能够在两个方向同时驱动 100MB。为便于实际网络设计，每个千兆位光纤通道管道，视销售商的实现方法，能够提供高达 100MBps 的持续吞吐量。一旦确定了这种可能性，那么，问题就是如何根据实际的带宽要求选择合适的 SAN 拓扑结构。

本书主要讨论 SAN 的关键技术和实际实现问题。SAN 已被用来解决磁带备份、数据挖掘、灾难恢复、服务器集群和其他超过传统服务器/存储器连接性的能力的应用。已经获得的实践经验将促进这些以及其他应用的可行解决方法的设计与实现。本书共分为四个方面内容，即技术基础：包括第 1、第 2 章；SAN 技术：包括第 3~7 章；SAN 应用：包括第 8、第 9 章；SAN 发展：包括第 10~12 章。

第 1 章“存储技术概要”将介绍当前各种存储概念、技术特点、用途和有关的存储设备类型。

第 2 章“高性能智能接口技术与外围总线技术”将介绍传统的智能技术和目前流行的和即将推广的智能接口技术以及外围总线技术。

第 3 章“存储器和联网概念”概述基本联网概念，确定 SAN 与传统结构和互补存储器解决方法（网络访问存储器：NAS）两者的关系。由于负责服务器和存储管理的 IT 专业人员不熟悉互联网概念，所以要注意对 SAN 结构所共有的这些联网原理。

第 4 章“光纤通道的内部结构”研究使 SAN 赋能的主要千兆位功能部件和协议。光纤通道是一种层次网络结构，提供高速物理传送的规则，提供这种高速传送与其他服务的高级应用协议之间的接口。由于技术内容缘故，本章对那些只想大概了解 SAN，不想深入研究字节级描述的读者可以跳过。

第 5 章“SAN 拓扑”讨论 SAN 实现时经常采用的三种光纤通道拓扑结构：点到点、仲裁环路和光纤。因为仲裁环路配置是最大存储网络安装基础，所以本章详细研究分析它的一些独特的协议要求。光纤开关提供增强特性，它与仲裁环路结合，可为 SAN 提供功能强大的设计方案。

第 6 章“SAN 中问题的隔离”简要介绍 SAN 提出的独特诊断挑战。为 SAN 环境选择的诊断工具是光纤通道分析仪。但分析仪价格昂贵，要有丰富的经验才能操作和解释。本章还提供了一些对诊断链路级问题和与协议有关的事件很有用的问题隔离技术。

第 7 章“SAN 的管理”介绍推荐给 SAN 互连使用的各种管理策略，以及文件、磁盘卷和备份的全局管理。服务器和磁盘阵列的传统管理方法大都限于通风、温度和电源状态

的管理。将联网放在服务器后面增加了复杂性，因此需要采用更加复杂的管理技术。在更高一级，把 SAN 设备的管理与文件管理和卷管理集成在一起，保证能完全控制 SAN 的保护应用。

第 8 章“光纤通道产品”介绍 SAN 的主要硬件和软件部件。适配器、控制器、开关和集线器在建立物理互连和维持合适的服务器/存储器对话传输协议方面发挥不同的作用。为适应市场的竞争性质，SAN 产品的功能性会因厂商而不同。此外，由于战略联合和因竞争而倒闭，厂商本身也在变化。因此，对专用 SAN 产品的讨论，在可能的地方，只限于那些对实现 SAN 最有用的功能部件。

第 9 章“应用研究”研究 SAN 技术的七种实际应用：全动感视频图像、预编辑操作、不占用 LAN 和不占用服务器资源的磁带备份、服务器集群、Internet 服务提供商、校园存储器网络和灾难恢复。其中每一种应用都给存储器联网带来了一些具体问题。任何网络的设计准则应该是应用驱动的。确定需要多少数据，从哪些资源和以什么样的速度得到这些数据，往往要比寻找合适的功能部件来满足这些需求更加困难。正如应用研究表明的那样，SAN 提供了一种能满足各种各样的要求，包括带宽、距离和冗余性等的灵活结构。

第 10 章“光纤通道的未来”讨论开发中的新能力。光纤通道是一种正在发展的千兆位技术。建议为混合速度环境提供 2GB 和 4GB 速度以及自动感知技术。光纤通道与 ATM 的互联是另一个感兴趣的方面，因为它允许宽域网在存储器网络之间进行高速通用数据存取。由于这些及其他创新将作为有用的产品来实现，所以 SAN 设计人员可广泛选择，以便满足企业日益增多的数据欲望。

第 11 章“InfiniBand 互联技术”介绍可能对存储网络产生巨大影响的一种新的串行总线技术。

第 12 章“主要厂商的 SAN 发展战略”介绍当前世界上主要的 SAN 产品供应商的产品战略；从中可看出未来存储技术的发展趋势。

在本书形成过程中，曾得到吴碧伟、李雯、杨烈文、周丽、陆晓亮、同志的支持和帮助，在此一并表示感谢。

由于受 SAN 技术复杂性和我们水平的限制，尽管我们在各方面均作了较大努力，但书中仍不免会有不妥之处，敬请读者批评指正。

作者

目 录

| | |
|---|----|
| 1 存储技术概要..... | 1 |
| 1.1 磁存储技术..... | 2 |
| 1.2 光学存储技术..... | 7 |
| 1.2.1 为什么要用光学存储设备..... | 7 |
| 1.2.2 光学存储设备的应用..... | 8 |
| 1.2.3 光盘的主要技术..... | 8 |
| 1.3 并行存储技术——RAID..... | 12 |
| 1.3.1 RAID 技术产生的背景及原因..... | 12 |
| 1.3.2 RAID 的基本概念..... | 12 |
| 1.3.3 实现 RAID 技术的主要方式与途径..... | 14 |
| 1.3.4 磁盘阵列技术主要研究方向..... | 14 |
| 1.4 NAS 技术..... | 14 |
| 1.5 SAN 技术..... | 15 |
| 1.5.1 SAN 与 NAS 的比较..... | 15 |
| 1.5.2 什么是 SAN..... | 16 |
| 1.5.3 SAN 架构几个明显的优点..... | 17 |
| 1.5.4 SAN 的前景..... | 18 |
| 1.6 备份技术..... | 19 |
| 1.6.1 备份的意义..... | 19 |
| 1.6.2 备份的误区..... | 19 |
| 1.6.3 备份的特点..... | 20 |
| 1.6.4 备份的方式..... | 20 |
| 1.6.5 备份系统硬件..... | 20 |
| 1.6.6 备份系统的软件..... | 21 |
| 1.6.7 备份系统的步骤..... | 21 |
| 1.7 小结..... | 21 |
| 2 高性能智能接口技术与外围总线技术..... | 23 |
| 2.1 传统的智能接口技术..... | 23 |
| 2.2 SCSI——智能接口总线的“长青树”..... | 24 |
| 2.2.1 SCSI/SCSI-2 接口标准..... | 25 |
| 2.2.2 Ultra SCSI 接口/Ultra2 SCSI 接口技术..... | 25 |
| 2.2.3 SCSI 技术的新发展..... | 26 |
| 2.3 SCSI—3 体系结构..... | 27 |

| | | |
|--------|---|----|
| 2.3.1 | SCSI—3 标准简介 | 27 |
| 2.3.2 | SCSI—3 体系结构模型 | 29 |
| 2.4 | Ultra 160 SCSI、Ultra320 SCSI 的发展概况..... | 36 |
| 2.4.1 | Ultra 160 SCSI | 36 |
| 2.4.2 | Ultra 320 SCSI..... | 36 |
| 2.5 | IEEE 1394 接口 | 40 |
| 2.5.1 | IEEE 1394 的特点 | 41 |
| 2.5.2 | IEEE 1394 的应用 | 41 |
| 2.5.3 | IEEE 1394 拓扑结构 | 41 |
| 2.5.4 | IEEE 1394 的寻址方式..... | 43 |
| 2.5.5 | IEEE 1394 体系结构 | 43 |
| 2.5.6 | IEEE 1394 仲裁时序 | 44 |
| 2.5.7 | IEEE 1394 传输机制 | 45 |
| 2.5.8 | IEEE 1394 的应用及前景 | 45 |
| 2.6 | 光纤通道 (FC) 接口 | 45 |
| 2.6.1 | FC 的特点..... | 46 |
| 2.6.2 | FC 的连接..... | 46 |
| 2.6.3 | FC 总线结构..... | 46 |
| 2.6.4 | FC 拓扑结构..... | 47 |
| 2.6.5 | 光纤通道仲裁环 (FC-AL) 寻址 | 48 |
| 2.7 | USB 接口 | 48 |
| 2.8 | SSA (串行存储体系结构) | 49 |
| 2.9 | ISA 和 EISA 总线 | 50 |
| 2.9.1 | ISA 总线 | 50 |
| 2.9.2 | EISA 总线..... | 50 |
| 2.10 | VME 总线..... | 51 |
| 2.11 | VESA 总线 | 52 |
| 2.12 | PCI 总线 | 53 |
| 2.13 | 未来 IO (Future I/O) | 55 |
| 2.14 | 下一代 I/O 总线结构 | 56 |
| 2.15 | UMA 总线 | 57 |
| 2.15.1 | UMA 总线的体系结构 | 57 |
| 2.15.2 | UMA 总线的特点 | 58 |
| 2.16 | EV6 总线 | 58 |
| 2.17 | 小结..... | 59 |
| 3 | 存储器与联网概念..... | 60 |
| 3.1 | 在服务器的前面联网..... | 61 |
| 3.1.1 | 串行传送..... | 63 |
| 3.1.2 | 访问方法..... | 63 |

| | | |
|-------|---------------------|-----|
| 3.1.3 | 寻址..... | 64 |
| 3.1.4 | 数据打包..... | 64 |
| 3.1.5 | 数据包路由选择..... | 64 |
| 3.1.6 | 上层协议支持..... | 64 |
| 3.2 | 传统 SCSI 总线结构..... | 64 |
| 3.3 | 网络连接的存储器 (NAS)..... | 66 |
| 3.4 | 在服务器的后面联网..... | 67 |
| 3.5 | 小结..... | 67 |
| 4 | 光纤通道的内部结构..... | 69 |
| 4.1 | 光纤通道的层..... | 69 |
| 4.2 | 千兆位传送..... | 69 |
| 4.3 | 物理层选择..... | 71 |
| 4.4 | 数据编码..... | 73 |
| 4.5 | 有序集合..... | 74 |
| 4.6 | 成帧协议..... | 75 |
| 4.7 | 服务类别..... | 76 |
| 4.8 | 流控制..... | 78 |
| 4.9 | 名字和寻址约定..... | 78 |
| 4.10 | 小结..... | 80 |
| 5 | SAN 拓扑..... | 82 |
| 5.1 | 点对点拓扑..... | 82 |
| 5.2 | 仲裁环路..... | 83 |
| 5.2.1 | 环路物理拓扑..... | 83 |
| 5.2.2 | 环路寻址..... | 85 |
| 5.2.3 | 环路初始化..... | 86 |
| 5.2.4 | 端口注册..... | 90 |
| 5.2.5 | 环路端口状态机..... | 90 |
| 5.2.6 | 仲裁..... | 91 |
| 5.2.7 | 仲裁环路的非广播性质..... | 92 |
| 5.2.8 | 仲裁环路设计的考虑因素..... | 93 |
| 5.3 | 光纤通道结构..... | 97 |
| 5.3.1 | 光纤注册..... | 100 |
| 5.3.2 | 简单名称服务器..... | 100 |
| 5.3.3 | 状态变化通告..... | 101 |
| 5.3.4 | 专用环路支持..... | 101 |
| 5.3.5 | 光纤开关分区..... | 102 |
| 5.4 | 建立扩展的 SAN..... | 103 |
| 5.5 | 小结..... | 104 |

| | | |
|-------|--------------------------------|-----|
| 6 | SAN 中的问题隔离..... | 105 |
| 6.1 | 简单的问题隔离技术..... | 105 |
| 6.2 | 光纤通道分析仪..... | 107 |
| 6.3 | 小结..... | 109 |
| 7 | SAN 的管理..... | 110 |
| 7.1 | 存储网络管理..... | 111 |
| 7.1.1 | 带内管理..... | 111 |
| 7.1.2 | 带外管理..... | 112 |
| 7.1.3 | SNMP..... | 113 |
| 7.1.4 | HTTP..... | 114 |
| 7.1.5 | 远程通信网络..... | 115 |
| 7.1.6 | 存储网络管理问题..... | 115 |
| 7.2 | 存储资源管理..... | 116 |
| 7.3 | 存储管理..... | 117 |
| 7.4 | 存储器、系统和企业管理集成..... | 117 |
| 7.5 | 小结..... | 118 |
| 8 | 光纤通道产品..... | 119 |
| 8.1 | G 位接口转换器 (GBIC) | 119 |
| 8.2 | 主机总线适配器..... | 120 |
| 8.3 | 光纤通道 RAID..... | 122 |
| 8.4 | 光纤通道 JBOD..... | 124 |
| 8.5 | 仲裁环路网络集线器..... | 125 |
| 8.5.1 | 仲裁环路的星形拓扑..... | 126 |
| 8.5.2 | 网络集线器结构..... | 126 |
| 8.5.3 | 非管理网络集线器..... | 129 |
| 8.5.4 | 管理网络集线器..... | 129 |
| 8.6 | 开关网络集线器..... | 131 |
| 8.7 | 光纤通道交叉开关 (Fabric Switch) | 131 |
| 8.8 | 光纤通道——SCSI 网桥 | 133 |
| 8.9 | SAN 软件产品..... | 134 |
| 8.10 | 小结..... | 136 |
| 9 | 应用研究..... | 138 |
| 9.1 | 全动视频..... | 138 |
| 9.2 | 预压缩操作..... | 140 |
| 9.3 | 不占用 LAN 和服务器资源的磁带备份..... | 141 |
| 9.4 | 服务器群..... | 144 |

| | | |
|--------|-----------------------|-----|
| 9.5 | 网际服务提供商..... | 145 |
| 9.6 | 校园存储网络..... | 146 |
| 9.7 | 故障恢复..... | 147 |
| 9.8 | 小结..... | 148 |
| 10 | 光纤通道的未来..... | 150 |
| 10.1 | 带宽..... | 150 |
| 10.2 | 宽域网的光纤通道..... | 151 |
| 10.3 | 在企业网络中共存..... | 151 |
| 10.4 | 互用性..... | 152 |
| 10.5 | SAN的一揽子解决..... | 152 |
| 10.6 | 小结..... | 153 |
| 11 | InfiniBand 互联技术 | 154 |
| 11.1 | 体系结构层次..... | 154 |
| 11.1.1 | 拓扑结构..... | 154 |
| 11.1.2 | 协议层..... | 154 |
| 11.1.3 | 部件..... | 156 |
| 11.1.4 | 主要特征..... | 157 |
| 11.2 | 结构和操作..... | 158 |
| 11.3 | 管理模型..... | 158 |
| 11.4 | 软件体系结构..... | 159 |
| 11.5 | IBA 的优势..... | 160 |
| 11.6 | 小结..... | 161 |
| 12 | 主要厂商的 SAN 发展战略..... | 162 |
| 12.1 | IBM 的战略..... | 162 |
| 12.2 | 康柏公司的战略..... | 163 |
| 12.3 | SUN 公司的战略..... | 165 |
| 12.4 | HP 公司的战略..... | 167 |
| 12.5 | EMC 公司的战略..... | 168 |
| 12.6 | StorageTek 公司的战略..... | 168 |
| 12.7 | 小结..... | 169 |

1 存储技术概要

在人类信息技术发展历程中，数字技术是一个划时代的里程碑。从第一代电子计算机的诞生至今，数字技术引发了三次浪潮。

1) 以处理技术为中心，以处理器的发展为核心动力，产生了计算机工业，特别是 PC 的普及，推广了数字技术的应用范围。

2) 以传输技术为中心，以网络技术的发展为核心动力，使人类社会进入到网络互连时代，人们无论何时何地都可以方便地获取和传递信息。

3) 正是由于上述原因，使人类社会产生了越来越多的数字化信息，信息量爆炸性地增长，引发了以存储技术为中心的新一轮数字技术浪潮，存储技术从幕后走到前台，跨入了一个崭新的时代。

随着现代电子技术、信息处理技术、通信技术的发展，人们越来越意识到计算机的重要性，高、精、尖的芯片及计算机系统不断出现并投入使用，VLSI 技术加速了处理器芯片与存储器芯片的发展。但由于外存设备固有的特性，使处理器、存储器、外存发展不平衡，产生了“存储器墙”、“I/O 瓶颈”等问题。

主机系统中性能冲突既反映在存储器系统与处理器的速度失配上，也反映在存储器系统的其他层面上的速度失配上，如 Cache-DRAM 接口，Cache 存储器速度急骤增长的同时，而 DRAM 增长不快。假设 Cache 访问时间等于一个指令周期，DRAM 访问时间是 Cache 的 1/4，Cache 失效是强制性的失效，这种失效约为 1%。又设 Cache 芯片性能以 CPU 的性能同样速率增长，为每年 80%，而 DRAM 以每年 7% 增长，DRAM 访问在 2000 年为 1.52CPU 周期，2005 年为 8.25CPU 周期，2010 年为 98.8CPU 周期，在这种情况下系统的性能不是由 CPU 速度决定的，而是由存储器的速度决定，即遇到了所谓的“存储器墙”。

“I/O 瓶颈”问题比“存储墙”更严重，已从巨型机领域扩散到通用计算机领域，服务器、工作站以及低档微机。人们对 I/O 的要求越来越高。以往人们研究巨型机 I/O 性能的技术如 I/O 的并行、通道、缓冲、总线等技术已逐渐应用到服务器、工作站以及低档微机等新领域中。20 世纪 80 年代以后，外存储新技术的涌现，产生了一批新型外存设备，而且应用面越来越宽广。

“I/O 瓶颈”产生的主要原因有：

1) CPU 速度在过去几十年内已有显著的增长，而 I/O 设备的速度由于受机械构件本身特性的限制仅以较低的速度增长，例如 CPU 速度以每年 50%~100% 增长，而磁盘访问时间大约每十年只减少 1/3，近期虽有提高，但总体趋势会保持基本不变。

2) 在并行分布共享的多机系统中加重了主机与 I/O 速度的失配性。

3) 多媒体、可视化及巨型复杂课题等一些新的应用领域产生了日益增长的 I/O 要求。“I/O 瓶颈”仍是海量科学计算的障碍。例如，许多科学计算应用需求运行产生 1GB 的 I/O 请求。通过一道地球科学程序，它需执行总 I/O 数据量为 28GB；而计算物理流体力学领域则总 I/O 需 1TB 数量级；目前二级存储 I/O 速率仅为数十 MB/s，正向 GB/s 发展，不能满足其要求，另一个应用领域是数据库，描述人类基因，结果涉及 GB 数据甚至 TB 数据的

大规模数据库检索，在图像、多媒体、可视化领域，如 HDTV、VOD 等并行 1 万个请求源，每个请求源速度为 0.5MB/s，则其流量 5000GB/s 等。

I/O 性能集中体现在 I/O 的速度，包括传输速率和响应速率上。为了达到这些要求，许多技术引入到现行的系统中，以改善存储系统包括 I/O 系统的性能。解决“I/O 瓶颈”的方案有：多端口（双端口）存储、多级存储模式、缓冲技术、交叉存储、通道技术、队列管理及外存调度策略等。近来对 I/O 的研究又产生了文件分条、RAID 技术、并行 I/O 技术、SAN 技术等。I/O 性能提高可通过以下途径解决：

1) 发展外存设备技术本身是奠定 I/O 性能提高的基础。无论是磁记录方式、记录密度、磁头技术，还是精密机械光电伺服系统等技术，都大大地提高外存的 I/O 速度、容量和可靠性能。

2) 改进提高 VLSI 技术来促进高性能智能接口技术、高性能外围总线技术的发展。这些技术是发挥 I/O 性能的重要硬件平台。

3) 并行处理技术引入到外存储系统中，出现了 RAID 技术、RAIT 技术、并行 I/O 技术等，从 I/O 系统的各个层面实现并行操作以提高 I/O 性能。

4) 标准化开放式的协议为系统平滑互联与系统集成提供了有效的连接方式，解决了系统接口的通用性、标准化、兼容性的问题。

传统的存储技术通常以大容量、高速度、低价格、小型化等特性来描述，而现代存储系统则要求具备高可靠性、高可用性、高性能（如 I/O 率与数传率）。系统的缩放性、易维护性和开放性等特点，它既是对现有存储技术的挑战，同时又刺激了各种存储新技术的发展。

人们已经认识到数据已成为人类最宝贵的财富，信息转变为数据，数据的价值越来越高。设备坏了可以更新，而数据丢失造成的损失是无法估量的，有时甚至是毁灭性的，因此信息存储系统的可维护性和可用性，数据备份和灾难性恢复能力往往是用户首先要考虑的。为了防止意外灾难等发生对数据的毁坏，关键数据还要考虑异地备份和容灾措施。同时人们也认识到目前计算机应用模式发生了重大变化，电子商务等网络服务要求全天候服务，也就是 7×24 小时乃至 365×24 小时服务，要求存储系统具有极高的可用性和快速的灾难恢复能力。集群系统、实时备份、灾难恢复都必须是为全天候服务开发的新技术。存储管理和维护必须实现自动化、智能化，存储系统同时具有多平台的互操作性和数据共享。

1.1 磁存储技术

计算机的主体包括中央处理器和主存储器，而这以外的部分则为计算机的外围设备或外围系统。外存技术的发展奠定了提高 I/O 性能的基础。

现代计算机系统中很重要的部分是计算机存储器，它分为主存和辅存。主存用来存放需立即使用的程序、数据，它要求存取速度快，通常由半导体存储器构成。辅助存储器作为主存的备份和补充，也可称为后备存储器、外存储器、二级存储器，它与主存不同，具有“非易失性”的特点，其容量大，成本低，通常断电后仍能保存信息。目前流行的外存贮器主要有磁表面存储器和光存储器等。

外存贮器的性能评价通常用如下技术指标来衡量：

- 存储密度；

- 存储容量；
- 寻址时间；
- 数据传输速率；
- 输入输出率（I/O 率）；
- 误码率；
- 价格等。

虽然磁记录原理未有根本性的变化，但是通过不断改进物理实体，包括磁头、记录介质和磁记录方式提高了外存 I/O 性能。20 世纪 80 年代以后，磁记录方式有了较大的发展，增添了新的编码技术并投入实际使用。常用的磁记录方式有：

- 归零制（RZ）；
- 不归零制（NRZ）；
- 改进不归零制（NRZI）；
- 调相制（PE/PM）；
- 调频制（FM）；
- 改进调频制（MFM）；
- 再改进调频制（M²FM）；
- 成组编码（GCR）；
- 游程长度受限码（RLL）等。

这些编码体制压缩了表征数据的有效编码长度，提高了数据利用率，同时也增加了容量。

外存设备的读写机构的特点决定了它的记录方式。磁带机是多道并行存取结构，一般采用调相制（PE/PM）和成组编码（GCD）。近年来高密度磁盘机则主要采用 RLL 码。

RLL 码（或 RLLC，Run Length Limited Code）称为游程长度受限码，是编码理论中研究码制变换、增强抗干扰能力而得出的一种编码。FM、PM 均可用 RLL 码描述。

RLL 编码的实质是将原始数据序列变换成“0”“1”受限制的记录序列。其编码规则是：

1) 先把输入信息序列变换“0”游程长度受限码，即任何两位相邻的“1”之间“0”的最大位数 k 和最小位数 d 均受到限制的新编码。

2) 用逢“1”变化不归零制方式进行调制和写入，具有自同步能力。

正确地设计 k 、 d 值，可以获得优良的性能。目前该码已广泛用于各种磁盘设备中。磁带机和软盘使用的成组编码记录 GCR（5，4）也是一种 RLL 码。GCR（Group Coded Recording）

码的编码规则是：把输入信息序列按 4 位长度分组，按表 1.1 给出的规则把 4 位信息变换成 5 位码字，最后再把编码用逢“1”变化不归零制方式调制。此同样具有自同步能力。用这种编码后，存储密度大大提高。

磁记录方式上为了抗干扰，采用了部分响应最大似然 PRML（Partial-Response Maximum Likelihood）处理技术。PRML 实际上是通信理论中部分响应 PR（Partial-Response）和最大似然（Maximum Likelihood）两种信号处理方法的统一运用。这两种技术所构成的 PRML 系统对抗干扰和提高记录密度起到了重要作用。它属于部分响应数字基带传输系统，其实现过程是：首先，进行码型变换；然后，进入部分响应信号通道；最后，经码型译码还原

为原始信息码。

表 1.1 GCR (5, 4) 变化规则表

| 信息组 D1 D2 D3 D4 | 码字 C1 C2 C3 C4 C5 |
|--------------------|----------------------|
| 0 0 0 0 | 1 1 0 0 1 |
| 0 0 0 1 | 1 1 0 1 1 |
| 0 0 1 0 | 1 0 0 1 0 |
| 0 0 1 1 | 1 0 0 1 1 |
| 0 1 0 0 | 1 1 1 0 1 |
| 0 1 0 1 | 1 0 1 0 1 |
| 0 1 1 0 | 1 0 1 1 0 |
| 0 1 1 1 | 1 0 1 1 1 |
| 1 0 0 0 | 1 1 0 1 0 |
| 1 0 0 1 | 0 1 0 0 1 |
| 1 0 1 0 | 0 1 0 1 0 |
| 1 0 1 1 | 0 1 0 1 1 |
| 1 1 0 0 | 1 1 1 1 0 |
| 1 1 0 1 | 0 1 1 0 1 |
| 1 1 1 0 | 0 1 1 1 0 |
| 1 1 1 1 | 0 1 1 1 1 |

PRML 系统的功能模块可以包括：可编程模拟滤波器、自适应横向滤波、A/D 转换器、Viterbi 译码器、GCR (8, 9) 编码器、时基产生器、扰码器解扰器以及预编译码电路等。

写通道时，NRZ 数据经过 GCR (8, 9) 编码器、预编码及写预补偿处理后，送写驱动电路，将数据记录在介质上。

读通道时，回读脉冲幅值并 A/D 转换前维持恒定，滤波器主要实现均衡，使信号精确整形，然后采样并将采样值送 Viterbi 译码器译码。其结果经过 GCR (8, 9) 译码和解扰，用 NRZ 信号调制，时基产生器要取得译码的同步时钟。

磁记录介质是指涂有薄层磁性材料的信息载体。它的底可分为软性介质（磁带或软盘片）和硬性介质（硬磁盘片）两种。而磁头是实现电磁转换的装置。目前磁记录介质与磁头的发展方向均需采用“双薄”技术，即薄膜介质与薄膜磁头。

薄膜头和薄膜介质技术促进了磁记录密度的发展。IBM 于 1994 年已利用新开发的磁阻薄膜磁头使硬盘驱动器存储密度高达 10^{10} B/in²，即每平方英寸可存储 100 亿个数据点，完全可与磁光记录相匹敌。虽然早已开发成功原感应式薄膜磁头，但这种薄膜磁头存在以下缺点：

- 1) 因具有很薄的主磁极端，其读出的波形的主峰两旁形成负峰（undershoot）。
- 2) 因线圈电阻较大而形成很高的热噪声。
- 3) 写入和读出头为一体，具有同一间隙密度，因此不能单独控制各自的最佳宽度。
- 4) 很难控制磁极端的磁畴结构，因而道宽的减少受到限制（理论值最小可达 $3\mu\text{m}$ ）。
- 5) 读出信号幅值决定于头盘界面之间的相对线速度，因此对小尺寸的驱动器盘片，内

道输出幅值太小。

为了克服上述缺点，满足高密度、小型化的要求，必须找到一种输出信号独立于头、盘的相对线速度的新型磁头。磁阻式磁头即属于此类。它对道宽没有限制，如 IBM $1\text{GB}/\text{in}^2$ ，则允许 $1\ \mu\text{m}$ 。它们均是利用 NiFe 薄膜材料的各向异性磁阻效应（AMR）而开成 MR。采用 NiFe/Ag/NiFe 巨磁阻效应材料可使记录密度达到 $10\sim 20\text{GB}/\text{in}^2$ ，实验室则可使记录密度达到 $100\text{GB}/\text{in}^2$ 。

MR 的基本工作原理是，在某种基体材料上沉积 $400\sim 600\text{\AA}$ 的具有磁各向异性的 NiFe 薄膜，用掩模光刻法、化学法或干法刻出一长方磁阻片（MRStripe），其长度与磁性膜易轴方向一致。然后沿此长度方向加一恒向电流（传感电流），同时在膜面内垂直于电流方向施加横磁场。当外磁场等于零时，磁阻片电阻率最大。当横向外加磁场逐渐减少，当横向磁场足够大时，磁阻片的磁化方向与外加磁场保持一致，其电阻率最小。通过磁阻头两端产生周期电压脉冲调节磁灵敏度。这一工作机制保证了 MR 的输出信号强度独立于头、盘界面之间的线速度，具有很高的信噪比，而且写入和读出磁头参数可各自调节到最佳值。

为了提高磁盘的容量和速度，通常可以采用如下技术：

1) 采用高密度磁头记录技术。MR（Magneto Resistive head）即磁阻式磁头的开发成功，是磁记录科技的一项重大突破。目前磁记录的面密度保持每年递增一倍的势头，已达到 2000 兆位每平方英寸。据理论估计，磁记录的面密度将达到 10000 亿位每平方英寸，线密度可实现每个原子对应一位。

2) 采用低信噪比的信号处理技术，其中包括信号记录的编码方式、读出过程信号的均衡和校验。目前 IBM 和日立均采用部分响应最大似然 PRML 技术，它也能够提高磁记录的容量。

3) 采用精密伺服定位技术，磁道定位技术由传统的磁信号伺服转变为如同光盘一样的光信号伺服。日立公司已达到了每英寸 17000 道的高磁道密度。目前还可达到更高的密度。

4) 采用高密度、低噪音的记录介质技术，如 Co，Pt，Cr 等介质。

5) 采用不断改进编码技术。

6) 采用高速磁盘控制技术，增加数传率的控制方案，增加超高速缓冲 Cache 的容量、串口传送方式，提高主轴转速。目前磁盘驱动器主轴转速可达 15000 转/分，实验室则可达 20000 转/分。

7) 采用高性能智能接口，如 SCSI-3、Ultra 160 SCSI、Ultra 320 SCSI、光纤通道等。

8) 采用并行处理技术，如磁盘阵列技术产生了 RAID、RAIT 等产品。

目前 GMR 磁头、加钉层介质、超低浮动技术和 PRML 信息处理技术等使记录密度有很大的提高。

目前，软盘技术采用了部分硬盘的实现技术，GB 级的软盘已面向用户。各种外存设备间技术渗透屡见不鲜，硬盘的伺服定位，扇区伺服、埋入伺服、光伺服等技术用于软盘驱动器，使得软盘容量有了飞跃；伺服磁道跟踪技术，1.7RLL 记录码，薄膜磁阻头用于盒带驱动器使得容量获得极大的增长，光盘中用浮动磁头进行磁场调制可实现读写；区位记录技术不仅普遍用于小型磁盘驱动器中，使容量有约 30% 的增加，它也适用于光盘驱动器及软盘；在磁盘驱动器中借助光伺服技术可大幅度提高道密度；WORM 型光盘技术磁带化，形成了 WORM 记录方式的光带存储设备；将浮动原理引用到软盘上，出现了 Bernoulli 软盘技术；作为解决高密度磁盘微弱信号处理的 Viterbi 与 PR 检测术在新型光盘中也同样会

得到非常的效果，利用磁盘阵列思想实现磁带阵列、光盘阵列等也不是没有可能。具有系统自动管理功能的磁带库，光盘库等也已在某些领域得到了推广应用。

磁带曾作为格式统一、互换性强的外存储器，深受广大用户的青睐。数据压缩技术在磁带存储设备中的应用较为广泛。目前常用的磁带技术如下：

1) DDS (Digital Data Storage)。数字式数据存储 (DDS) 是基于数字式音频磁带 (DAT) 格式而发展的。在用 DAT 记录音乐和声音时，DDS (也属于 4mm 磁带) 用于记录数据。但记录音乐的 4mm 磁带的使用已经减少了，只占目前磁带存储市场的 1%。

DDS 是 Hewlett-Packard (HP) 和 Sony 于 1988 年根据潜在的用户而开发的，DDS 的制造集团现在超过 20 个商家。DDS 是根据基本的 DAT 格式扩展而设计的，加入了误差修正和其他特点，以便适合于数据存储的格式。

DDS 是数字式的，这表示它使用了两个电性状态：无或有 (Nothing or Something)。当达到了信号的阈值时，就记录到两个状态中的一个。数字记录优于模拟记录的地方是，即使信号的强度有很大的变化，只要是超过了信号阈值，就将检测到 0 或 1。

2) DAT (Digital Audio Tape)。数码音频磁带技术 (DAT) 最初是由 HP 和 Sony 共同开发的。这种技术以螺旋扫描记录 (Hellical Scan Recording) 为基础，将信号转化为数字后再存储下来。早期的 DAT 技术主要应用于声音的记录。随着技术的不断完善，DAT 又被应用于用户系统或局域网，它以非常合理的价位提供高质量的数据保护。

在信息存储领域里，DAT 一直是广泛应用的技术之一，而且种种迹象表明，DAT 的优势还将继续保持。这种技术之所以大受欢迎，还由于它具有很高的性能价格比。首先，在性能方面，这种技术生产出的磁带机平均无故障工作时间长达 20 万小时 (新产品已达到 30 万小时)；在可靠性方面，它所具有的即写即读功能，能在数据被写入之后马上进行检测，不仅确保了数据的可靠性，而且还节省了大量时间。其次，磁带机种类繁多，能够满足绝大多数网络系统备份的需要。对于网络系统，DAT 产品由于已被证明的高兼容性、高可靠性与介质的低成本而被广泛接受并应用。第三，硬件数据压缩功能大大加快了备份速度，而且压缩后的数据安全性更高。第四，由于该技术在全世界都被广泛应用，因此在全世界都可以得到该技术产品的持续供货与良好的售后服务。

3) DLT (Digital Linear Tape)。数码线型磁带技术 (DLT) 最早于 1985 年由 DEC 公司开发，主要应用于 VAX 系统。尽管这种技术性能出众，但是由于价格昂贵，因而在 1993 年销售量曾降到最低点。后来随着高档服务器的存储容量超过其他磁带机技术所能提供的容量，DLT 又重新成为信息存储领域里的热门技术。

DLT 技术采用单轴 1/2 英寸磁带仓，以纵向曲线记录法为基础。DLT 产品定位于中、高端的服务器市场和磁带库应用系统。目前 DLT 驱动器容量为 10~35GB，数据传送速度为 1.25~5MB/s，如果硬件采用数据压缩，则可以使容量与速度增加一倍。但 DLT 技术也存在相应的劣势，如驱动器与介质造价昂贵；主系统与网络之间通道狭窄，限制了传送速度，增加了操作时间；非标准的外形设计是内部设置受到很大限制；单一的产品提供使渠道受限。总的来说，DLT 人被认为是未来型产品，发展很快，但目前只会被少数需要高性能备份的系统用户所使用。

4) LTO (Linear Tape Open)。LTO 即线性磁带开放协议，是由 HP、IBM、Seagate 三家厂商在 1997 年 11 月联合制定的。它结合了线性多通道、双向磁带格式的优点，基于服务系统、硬件数据压缩、优化的磁道面和高效率纠错技术，来提高磁带的性能。