

中小企业 存储技术 初探

浦云明 (厦门集美大学信息工程学院 361021)

杨健 (厦门象屿码头有限公司 361006)

摘要: 本文结合厦门象屿码头有限公司的存储方案,初步探讨了中小企业采用存储技术实现数据的高可靠性和高可用性,涉及磁盘阵列、数据备份和 SAN 技术。

关键词: RAID 存储 备份



1 引言

随着中小企业信息化的推进,为满足客户对数据按 24 x 7 标准进行访问,企业对数据的可用性提出了前所未有的高标准,安全、可靠的存储技术和方案是每个信息主管必须面对的问题。近年来存储技术在厂家和用户需求的推动下得到了长足发展,特别是在金融、证券、因特网的应用方面。随着数据不断膨胀及海量数据的出现,存储技术和方案得到了广泛应用,包括:磁带备份、磁盘阵列、SAN 技术等,存储主流厂商 EMC 的存储系统、康柏的 SAN 方案、IBM 的企业存储方案等。虽然新的技术和方案更可靠安全,但就中小企业而言,在追求数据高可用性的同时,更多需要考虑实施存储方案的投入/产出比。笔者认为磁带备份和磁盘阵列(包括双机容错)还是中小企业存储方案的首选。

2 磁盘阵列 RAID 技术

RAID 是廉价冗余磁盘阵列 (Redundant Array of Inexpensive Disks)。1987 年 12 月加州伯克利分校的 David Patterson 等人,首次发表了 RAID 技术的论文,以解决小磁盘组的 MTBF (平均故障间隔时间) 问题,主要论点包括一组物理磁盘作为一个逻辑磁盘使用,用户数据根据定义的规则分布在这一组物理磁盘上,引入冗余磁盘保证一个磁盘故障后能恢复数据。并提出了 RAID1 到 RAID5 共 5 个级别,后来工业界补充了 RAID0 和 RAID6。磁盘阵列将同一阵列的多个磁盘作为一个逻辑盘,数据以分段的形式存放在磁盘阵列上,分段是把数据分为小段跨越分布在各物理磁盘上。

RAID 0, 又称磁盘分段 (Disk Striping), 数据段按顺序分布在各物理磁盘上,无容错能力,单个磁盘故障可导致系统故障(磁盘数一般不超过 4),无冗余;如四个物理磁盘的 RAID 0, 数据 ABCDEFGH 的分布如图 1, 每一字符表示一数据段:

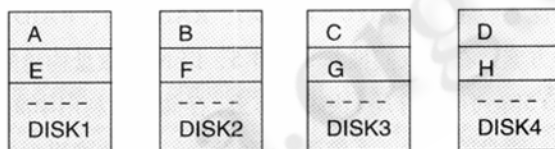


图 1 RAID 0 数据分布

RAID 1 即磁盘镜像 (Disk Mirroring), 是两块硬盘数据完全镜像,安全性好,有容错能力,允许单个磁盘故障,但空间浪费大。为改进性能,采用 2 对或更多对磁盘,数据在各磁盘对上分布存储,并在每对磁盘上作镜像,称为 RAID 0+1。RAID 0+1 具有最好性能和最高容错能力。如 2 对共 4 个磁盘作镜像如图 2。

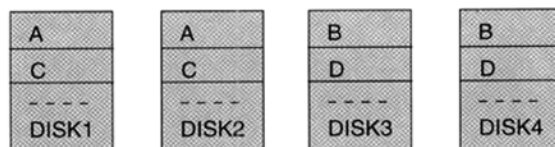


图 2 RAID 0+1 数据分布

RAID 2, 又称并行阵列, 采用共轴同部技术和海明码纠错, 存取数据时, 整个磁盘阵列一起动作, 在各个磁盘的相同位置作平行存取, 可并行传输/输出, 允许单个磁盘故障; RAID 3, 类似 RAID2, 但采用 BIT WISE 的专用奇偶校验盘。

RAID 5 磁盘分段加奇偶校验, 校验数据循环分布在各个磁盘, 允许单个磁盘故障, 校验数据是同一位置的数据异或 (XOR) 运算得到, $P(A,B,C)=A \text{ XOR } B \text{ XOR } C$, 见图 3。RAID 4 与 RAID5 类似, 但校验数据使用专用磁盘; RAID 6 类似 RAID 5, 使用 Reed-Solomon 码生成 2 组校验数据循环存放在 2 个磁盘上, 允许 2 个磁盘故障, RAID 6 的性能最好, 但实现复杂且成本高。

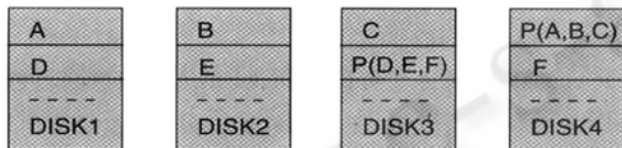


图 3 RAID 5 数据分布

目前常用的 RAID 磁盘一般使用 SCSI 磁盘, 对于企业来说一般考虑 RAID 0、RAID 1、RAID 0+1、RAID 5 四种方案。在此我们推荐 RAID 5, 以 N 块磁盘构建的 RAID 5 阵列可以有 N-1 块磁盘容量, 存储空间利用率高, 任何一磁盘的故障, 系统还能正常运行, 当故障磁盘修复或更换后, 故障磁盘的数据经校验计算而恢复, COMPAQ 公司也建议对不超过 14 个物理磁盘的系统使用 RAID 5。

3 磁带备份技术

磁带备份技术是传统的数据安全和存储技术, 包括备份软件和磁带设备, 经备份软件将用户数据从数据库备份到磁带。当系统故障时, 新系统或修复系统将数据从磁带恢复到数据库中, 对非实时容错的双机方案, 通过磁带备份可保持数据库准同步。例如民航计算机订座系统, 一台备份主机位于北京三里屯, 一台主机位于北京东四, 两台主机数据库通过磁带实施冷备份, 东四主机每天数次磁带备份数据库, 并运往三里屯装入备份主机数据库。备份技术发展到如今, 磁带备份技术还是许多公司的安全措施之一, 备份软件有 CA 公司的 ARCserve、VERITAS 公司的 Backup Exec 等。

4 SAN 存储技术

存储区域网络(SAN)是一个专有的、集中管理的信息基础架构 (e-infrastructure), 支持存储和服务器间的任意连接。SAN 解决方案包括软件、服务器、SAN 光纤组件和存储设备, 以实现存储整合和数据共享, 改进备份与恢复过程。SAN 是以光纤结构为基础的一个开放标准, 是新一代存储的发展方向 and 架构。EMC 和 COMPAQ 等主流厂商已推出了相应的解决方案。

4.1 EMC 的存储方案

EMC 公司作为存储设备主流厂商, 有一流的存储技术和产品, 如 Symmetrix 企业存储系统、SRDF 远程镜像软件、Timefinder 文件备份软件、Fastrax 数据传送平台等产品。Symmetrix 5000 企业存储系统是一开放存储解决方案, 可提供 TB 级的存储容量, 供多台主机使用并被集中管理, 不需要专门的主机设备驱动程序, 因此当升级操作系统和服务器时不会影响业务运行。系统的 Symmetrix ESP 软件允许 UNIX、Windows NT、AS/400 和大型机信息存放在同一 Symmetrix 5000 系统中, 并被各主机系统并发访问, 是一个真正开放的企业存储系统。

4.2 COMPAQ 的存储方案

COMPAQ 推出了 RA4000/RA4100 SAN 解决方案, 可将多个服务器的存储器归并到一个 SAN 上, 包括主存储器和辅存储器, 利用 Insight manager 软件在一个位置集中管理存储器。使用热插拔硬盘、双路控制器、在线扩容等技术, 可以不停机地扩展 RA4100 SAN 解决方案。采用康柏 FC-AL 交换机, 可以在运行中向 SAN 添加硬盘、阵列和额外的附件。每个 FC-AL 交换机可以支持 6 个服务器、9 个 RA4000/RA4100 存储器附件或 3 个光纤通道磁带控制器。

存储系统使用 Ultra3 阵列控制卡 5300 系列, 它有全新内存体系结构和 RAID 技术, 单个 PCI 插槽所带硬盘容量可扩充到 2.4TB, 每台服务器的存储总量可扩展到 14.4TB。增加了每通道 160MB/s 的 Ultra3 SCSI 通道, Ultra2 和 Ultra3 的混合匹配保证各代数据的兼容性。存储阵列的高速缓存 (32/64/128MB 可选) 具有后备电池设计, 一旦服务器或控制卡发生故障, 可保护高速缓存的数据。

5 厦门象屿码头数据存储方案

厦门象屿码头是一个集装箱专用码头, 历年的生产数据通过 CTMS 集装箱管理系统和 ORACLE 数据库存储, 因此必须保证数据的安全, 否则将给公司带来不可估量的损

失。系统设计之初就考虑了磁带备份和磁盘阵列2种方法,采用主从服务器结构,见图4。

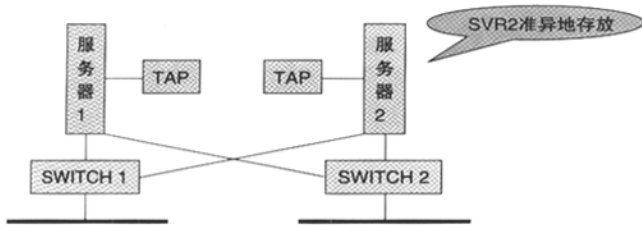


图4 象屿码头主从服务器结构图

5.1 备份

备份数据库用两种方法实现：磁带备份和网上数据备份，磁带备份采用了Veritas的Backup Exec备份软件，HP12000磁带库，一天一次热备份（凌晨1:00），一周一次冷备份（数据库shutdown），磁带实行异地存放。网上数据备份使用ORACLE数据库提供的Import/Export功能实现数据由主服务器到备份服务器的转移，一天一次（中午12:00进行），以上两种备份的结合，最坏的情况是可能丢失半天数据。

5.2 磁盘阵列

服务器是COMPAQ设备,采用FAST SCSI磁盘阵列,设计之初使用了5个4.3GB硬盘,实施RAID 0+1,其中2对(4个)磁盘作镜像,数据在各磁盘对上分布存储,并在每对磁盘上作镜像,保证了2套数据在一对磁盘上,另一磁盘作为系统盘不参加阵列。在NT升级时,为扩大存储容量,在分析数据可靠性和可用性后,磁盘阵列实施RAID 5,因此实际容量为 $(5-1) * 4.3GB$ 。从使用情况看,早期的RAID0+1和现在的RAID5设计是比较理想的。

5.3 象屿码头实时双机容错磁盘阵列方案

虽然现有主/从服务器实施了磁盘阵列和数据备份措施,但不能自动切换,一旦主服务器发生故障,仍需要人工切换。由于业务扩展,码头将对服务器升级,因此将采用最新容错技术,通过一条不占用网络带宽的专用高速链路和一套使用独立I/O请求的驱动程度级别的镜像设计,不但确保NT服务器的数据获得传统意义上的双服务器热备份,更确保其他重要网络资源(应用程序、文件系统和打印机等)获得同等程度的高可靠性。这样双服务器可实现“ACTIVE TO ACTIVE”的同等级别并互为备份的容错方式:在双方互相而持续地监控镜像资源的过程中,如果其中一台由于硬件或软件原因发生故障失效,另外一台可

在保证提供自己原有服务的同时,启动失效服务器的应用程序、文件系统、IP地址和打印机等网络资源服务从而取代其服务器功能。以上切换可由用户根据环境要求和硬件设备能力自行定义,切换时间仅以秒计。

此方案以支持集群的SCSI Channel存储子系统为基础。外置磁盘阵列柜,内装10个9G硬盘分两组,5磁盘个1组,采用RAID 5,两组磁盘相互镜像(Mirroring)。双机容错方案见图5。

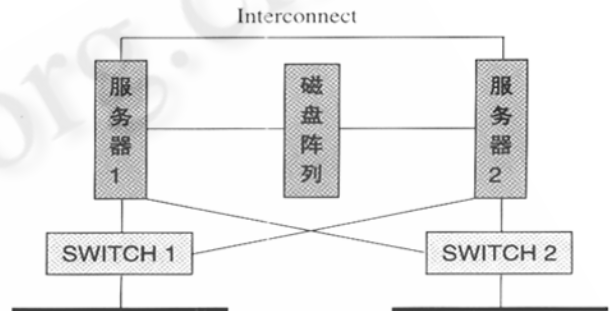


图5 象屿码头双机容错方案原理图

参考文献

- 1 华艺集团(中国)容错电脑系统公司, Oaraid 磁盘阵列系统, 1998.
- 2 Pete McLean, An Introduction to RAID Redundant Arrays of Inexpensive Disks, DEC, 24 April 1991.
- 3 COMPAQ 公司, 康柏存储产品销售大全, 第二版, 2000年8月.
- 4 民航计算机订座系统的备份, 计算机系统应用, 2000.10.

