

文章编号: 1005—8893 (2001) 02—0043—03

硬件磁盘阵列的使用^{*}

马正华¹, 吴 斌², 朱正伟¹

(1. 江苏石油化工学院 计算机科学与工程系, 江苏 常州 213016; 2 江苏常州数控机床厂 信息中心, 江苏 常州 213012)

摘要: RAID (廉价磁盘冗余阵列) 技术是提高存储子系统性能和可靠性的一种有效手段。RAID 可以由软件或硬件实现, 硬件 RAID 又可以分为主机内部硬件 RAID 和主机外部硬件 RAID。硬件 RAID 不依赖操作系统, 具有比软件 RAID 更高的运行效率和容错能力。主机内部硬件 RAID 又比主机外部硬件 RAID 具有更高的运行效率和管理能力。主机内部硬件 RAID 广泛应用于高端网络服务器中。以 Mylex DAC960 PCI RAID 适配器 (主机内部硬件 RAID 控制器) 为例, 对主机内部硬件 RAID 的基本概念、安装和配置、以及故障后的重建进行介绍。

关键词: 廉价磁盘冗余阵列; 硬件 RAID; 主机内部硬件 RAID; 存储系统; 网络服务器

中图分类号: TP 333. 3

文献标识码: A

1 RAID 及 RAID 适配器

RAID 是廉价磁盘冗余阵列的缩写, 其基本思想是使用容量较小的较为经济的硬盘组合在一起以代替大容量的单个硬盘并通过使用冗余的硬盘存放数据校验信息, 提高存储系统的可靠性。RAID 技术是提高存储子系统性能和可靠性的一种有效手段。

RAID 可以由软件或硬件实现^[1]。软件 RAID 的阵列操作和管理功能由运行于主机 CPU 上的阵列软件控制。阵列软件与其它同时运行的应用程序争夺 CPU 时间和内存, 软件 RAID 降低了服务器整体性能。硬件 RAID 由硬件—磁盘阵列适配器实现。磁盘阵列适配器上有独立 CPU, 具有独立于主机的阵列控制系统, 不占用主系统资源。硬件 RAID 的阵列操作和管理可以离开主机 CPU, 在阵列硬件自身的处理器上进行。当阵列硬件的处理器执行阵列功能时, 主机 CPU 可以执行用户的其它应用程序。硬件 RAID 通过阵列控制硬件本身独立的处理能力处理阵列操作, 减少服务器处理阵列的时间, 提供比软件 RAID 更高的运行效率。硬件

RAID 不依赖操作系统, 一个规划良好的硬件 RAID, 在操作系统发生故障时, 硬件 RAID 不会因为操作系统的崩溃而被破坏, 提供比软件 RAID 更高的容错能力。先进的 RAID 技术包含在线的硬盘热交换, 故障硬盘的在线重建恢复。这些优异的性能大大减少主机停机时间。

硬件 RAID 可以分为主机内部硬件 RAID 和主机外部硬件 RAID^[2]。主机内部硬件 RAID 安装在服务器主机“内部”, 直接连接到 I/O 总线 (如: EISA 总线或 PCI 总线)。主机外部硬件 RAID 安装在服务器主机“外部”, 典型的在一个外部阵列包中, 外部阵列包连接到服务器的一个 SCSI 适配器上。主机内部硬件 RAID 安装在主机上, 存储管理软件能够直接与阵列控制器通讯以便在局域网中的任何客户端对阵列进行有效管理和监控。尽管大多数主机外部硬件 RAID 提供多个 SCSI 通道用于配置阵列, 然而它们只可用单个 SCSI 通道 (带宽 20 MB/s) 连接到主机上。这产生了一个重大的性能瓶颈。主机内部硬件 RAID 直接连接到主机的 I/O 总线, 在 PCI 总线上可以以高达 132 MB/s 的速度传输数据。主机内部硬件 RAID 比主机外部硬件 RAID 具有更高的运行效率。尽管如此, 主机外部

* 收稿日期: 2001—04—09

作者简介: 马正华 (1962—), 男, 江苏昆山人, 副教授, 主要从事计算机技术及应用开发研究。

硬件 RAID 可以构造特大容量存储系统, 也可以构造双机容错系统, 实现更高档次的容错系统。硬件 RAID 因为其高性能而广泛应用于网络服务器中。

硬件 RAID 有许多容易混淆的概念, 本文以在 HP 网络服务器 HP LxPro 上安装、使用 Mylex DAC960 PCI RAID 磁盘阵列适配器为例, 对常用的主机内部硬件 RAID 的主要的概念和使用进行介绍, 以便读者对硬件 RAID 有更加清晰的认识和理解。

HP LxPro 是 HP 公司的高端网络服务器产品, 采用 Intel Peetium PRO200 CPU^[3]。

Mylex DAC960 PCI RAID 适配器^[4~6]是 Mylex 用于实现硬件 RAID 的高性能阵列控制器。DAC960 可以支持工业标准 RAID0, 1, 5 等 RAID 级别。DAC960 有 1~3 个完全独立的 SCSI 通道, 可以连接 SCSI 硬盘、CD-ROM 和磁带机。对于内部硬件 RAID 使用适配器的内部 SCSI 通道, 对于外部硬件 RAID 使用适配器的外部 SCSI 通道。DAC960 使用 32 位精简指令微处理器, ASIC 逻辑阵列和读写缓存减少 CPU 负荷, 增加磁盘 I/O 能力。DAC960 3.x 固件支持每通道 15 个硬盘, 每控制器可支持 32 个系统盘, 并具备坏数据管理, 磁盘上保存配置, 阵列自动重建, 自动扇区映射, 自动失效硬盘检测等特殊功能。

2 实现 RAID 的基本步骤

使用 DAC960 RAID 适配器实现硬件阵列的基本步骤如下:

(1) 阵列的规划

在安装磁盘阵列前应当根据具体对存储系统的需求, 制订阵列规划。

假设要安装 WindowsNT Server 和 Sybase SQL Server, 希望将系统程序安排在具有 RAID-1 级别(镜像)的磁盘上, 数据安排在 RAID-5 的磁盘上。制订阵列计划如下:

RAID-1 级别只使用 2 块物理磁盘; RAID-5 级别至少需要 3 块物理硬盘, 可以根据对实际数据容量的需求确定使用物理磁盘的数量。计算方法: RAID-5 容量 = (参与 RAID-5 的物理硬盘数量 - 1) × 单个物理硬盘容量。如果数据需要 20G 空间, 参与 RAID-5 的物理硬盘数量 = 6 (个), 考虑到使用 1 块硬盘作为备用硬盘。总计需要 9 块硬盘 (见图 1)。

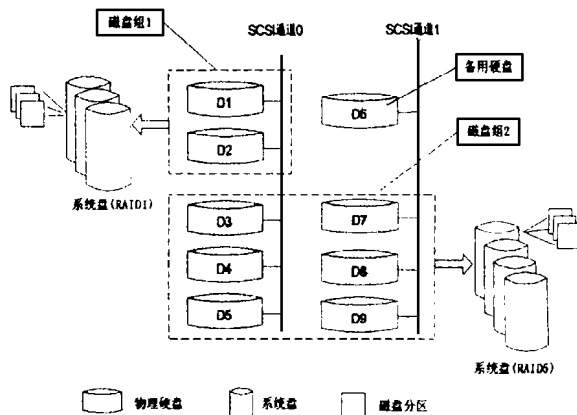


图 1 硬件 RAID 示意图

(2) 关闭所有电源。

(3) 将 DAC960 PCI 适配器插入主机的 PCI 插槽内。应当注意主机各个 PCI 插槽引导顺序可能有优先级区分。参考主机资料选择合适的 PCI 插槽。

(4) 将 SCSI 硬盘连接到适配器上。在连接前应当保证每个 SCSI 通道上的硬盘的 SCSI 地址不重复, 也不与控制器本身的 SCSI 地址重复, 可以通过 SCSI 硬盘的跳线改变 SCSI 地址。为保证磁盘阵列的性能, 各个硬盘尽量均匀分布在多个 SCSI 通道上, 同时尽量不要将 SCSI 硬盘与光驱或磁带设备连接到同一个 SCSI 通道上 (如果难以避免可以考虑使用主板上的 SCSI 通道或增加额外的 SCSI 适配器用于连接光驱和磁带设备)。应确保每个 SCSI 通道必须正确终结, 并保证在任何硬盘的移去不会改变 SCSI 终结状态。

(5) 打开电源, 用 DOS 启动盘启动主机。

(6) 运行阵列配置工具 DACCF 建立磁盘阵列, 建立可以安装网络操作系统的系统盘。

i 低级格式化各个物理盘

一般 SCSI 硬盘出厂时, 已经完成低级格式化, 所以一般可以略过此步骤。但是在硬盘失效后, 如果怀疑硬盘有物理问题或缺陷, 可以重新做低级格式化。

ii 定义一个或多个物理磁盘组 (PACK)

磁盘组是多个物理硬盘的组合。磁盘组中硬盘的个数决定了建立在磁盘组上的系统盘允许的 RAID 级别。如 3 个硬盘组成的磁盘组, 可以建立 RAID-0 或 RAID-5 的系统盘, 但是不能建立 RAID-1 的系统盘; 而 2 个硬盘组成的磁盘组, 只能建立 RAID-1 的系统盘。

如果希望将某块硬盘作为故障后的备用硬盘,

不要将它定义在任何一个磁盘组中。将来如果某块硬盘故障, 可以使用该盘接替故障硬盘, 以便在线修复, 保证系统正常工作且容错系统快速重建。

iii 在每一个物理磁盘组上建立有确定 RAID 级别的一个或多个系统盘 (System Disk)

系统盘建立在磁盘组上并有确定 RAID 级别, 它是磁盘组的部分或全部, 一个磁盘组可以建立一个或多个系统盘。系统盘可以建立的 RAID 级别由磁盘组中硬盘个数确定。一个磁盘组上各个系统盘的 RAID 级别可以是不同的。

iv 初始化各个系统盘

格式化好的系统盘, 从操作系统的角度, 相当于我们微机的多个硬盘, 只是这里的系统盘底层可能已经做了容错, 实际物理硬盘的特性被磁盘阵列硬件所封装。

(7) 安装操作系统

操作系统的安装可参照操作系统的资料。底层的 RAID 对操作系统而言是透明的, 操作系统将前面建立的系统盘视作“物理盘”。在安装操作系统中可以根据需要在系统盘上建立磁盘分区。操作系统的安装与一般无容错系统的安装相同, 本文不再赘述。

物理硬盘、磁盘组、系统盘, 磁盘分区之间的关系如图 1 所示。

3 硬盘失效及 RAID 重建

在硬盘发生故障时, DAC960 会自动停止该物理硬盘工作, 并发出警告提示。因为 RAID 的存在, 系统仍然可以工作不产生任何影响, 只是系统此时已经不再能够写入校验信息了。这种状态应当尽早修复, 以便容错系统继续工作。这时可以更换失效硬盘, 然后使用阵列配置工具 DACCF 重建阵列, 从现有磁盘读出信息, 并在更换的硬盘上自动重新建立整个 RAID。如果在系统上已经有备用磁盘, 这项工作相当简单, 直接重建 RAID 即可。

参考文献:

- [1] CCG Europe B V Co. Hardware vs. Software RAID [EB/OL]. <http://www.ccgeurope.nl> 2000-09.
- [2] CCG Europe B V Co. Host-Based vs. SCSI-to-SCSI RAID [EB/OL]. <http://www.ccgeurope.nl> 2000-09.
- [3] Hewlett-Packard Co. Network Server Division. HP NetServer Lx Pro series UserGuide [M]. California: Hewlett-Packard Company, 1997. 6.
- [4] Hewlett-Packard Co. Network Server Division. HP NetRaid Installation and Configuration Guide [M]. California: Hewlett-Packard Company, 1997. 6.
- [5] Mylex Corporation. PCI RAID Controller Installation Guide (DAC960) [M]. California: Mylex Corporation, 1997. 6.
- [6] Mylex Corporation. Mylex Disk Array Controller Configuration Utilities Installation Guide & User Manual [M]. California: Mylex Corporation, 1998. 2.

The Usage of Hardware RAID

MA Zheng-hua¹, WU Bin², ZHU Zheng-wei¹

(1. Department of Computer Science and Engineering, Jiangsu Institute of Petrochemical Technology, Changzhou 213016, China; 2. The Center of Information, Changzhou Machine Tool Works, Changzhou 213012, China)

Abstract: RAID (Inexpensive Disk Redundancy Array) technology is an effective means to enhance performance and reliability of storage subsystem. RAID can be implemented with software or hardware, and hardware RAID can also be divided into host-based RAID and SCSI-SCSI RAID. Hardware RAID is not dependent upon operating system, having higher operating efficiency and fault-tolerant capability than software RAID. Host-based RAID has higher operation efficiency and better management capability than SCSI-to-SCSI RAID. Host-based RAID is widely applied to high-end network server. This paper uses Mylex DAC 960 PCI RAID adapter (controller of host-based RAID) as an example to introduce the basic concept of host-based RAID and its installation and configuration as well as its rebuilding after failure.

Key words: RAID; hardware RAID; Host-based RAID; storage subsystem; network server