双机双控容错系统的设计

欧阳珣 （华中理工大学 计算机学院 430074）

一、引言

近年来，随着计算机技术的飞速发展，服务器的性能有大幅度提高，服务器作为关键性事务 的业务主机已经成为可能。对于要求有高可用性和高安全性的系统，比如银行系统，用户提出了容错的要求，为此根据市场的需要，我们推出了双机双控容错系统。用两台服务器共同工作，当一台服务器的系统出现故障时，另一台服务器可确保系统工作正常运行，从而将系统风险降低到最低程度，保障了系统的高可靠性，高安全性和高可用性。

二、用户系统的要求

银行系统是我国生产环节中的重要组成部分，庞大的业务网络为全国城乡的发展做出了巨大的贡献。各个专业用户都根据各自的特点建立了各自的网络系统，其范围覆盖全国。这些重要的部门对数据的管理和传输就提出更高的需要。为此方案设计要满足和考虑以下的用户要求：

1.要能够支持现有的数据库（例如 Informix 和 Sybase）；

2.新方案能够保证数据不丢失（包括人为错误操作）；

3.由于用户数据量巨大，必须有大容量的磁盘来保存数据；

4.根据各地市用户不同的日交易量设计相应的方案；

5.针对有些用户已有设备加以改造，使之具有新方案所具有的功能。并具有高度容错能力（NetRAID）且速度要快。

三、双机双控容错系统的设计方案

针对不同的用户，有不同的方案选择，现以银行为例说明双机双控容错系统的方案设计，我们结合 Cluster 技术，选用 HP Netserver 和 HP 一起开发设计了双机双控容错系统，（分别针对三种不同交易量大小）并制定出了三个方案，如表 1 所示。这些系统还可以根据需要进行扩充。

四、双机双控容错系统的工作原理

1.硬件设置

HP 双机系统技术基础为近年来成熟起来的 Cluster 集群技术。即一组相互独立的服务器在网络中表现为单一的系统，并以单一系统的模式加以管理。在大多数模式下集群中所有的计算机拥有一个共同的名称，集群内任一系统上运行的服务可被所有的网络用户所使用。一个 Cluster 包含多台（至少两台）拥有共享数据存储空间的服务器。任何一台服务器运行一个应用时，应用数据被存储在共享的数据空间内。每台服务器的操作系统和应用程序文件存储在其各自的本地储存空间上。

Cluster 内各节点服务器通过一内部局域网相互通信。当一台节点服务器发生故障时，这台服务器上所运行的应用程序将在另一节点服务器上被自动接管；当一个应用服务发生故障时，应用服务将重新启动或被另一台服务器接管。系统硬件结构示意图如图 1 所示。

HP 双机系统为二台 HP NetServer 服务器，每台服务器拥有各自的系统盘，用来安装系统软件、数据库软件、应用软件和双机软件。两台服务器还拥有一共享的数据盘，用来存储应用数据。系统盘做 RAIDI 镜像冗余，数据盘做 RAID 5 或 RAID 50 级冗余。两台服务器拥有各自的 RAID 控制卡，系统为双控结构。

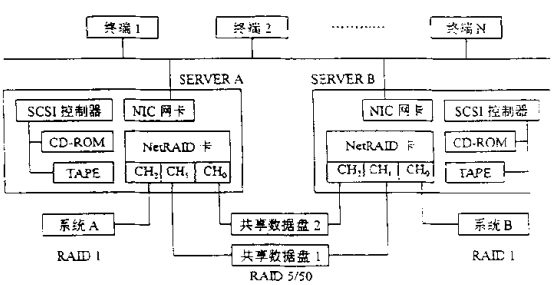

图 1 双机双控容错系统硬件结构示意图

表 1

适用环境		方案一	方案二	方案三
	银行级别	省级分行业务中心	地市级分行业务中心	县级二行业务中心
	日处理能力	10 万笔相关业务以上	5 到 10 万笔相关业务	5 万笔相关业务以下
硬件配置	CPU	Pentium Pro 200MHz × 4	Pentium Pro 200MHz × 2	Pentium 2 266、300 或 333MH
	内存	1GDIMM	512Mb DIMM	256M6 DIMM
	RAID Controller	NetRAID 1 通道×1 控制系统盘 NetRAID 3 通道×1 控制共享数据盘	NetRAID 3 通道×1 控制共享数据盘和系统盘	NetRAID 3 通道×1 控制共享数据盘和系统盘
	Rack Storage	Rack Storage 8×3 一台用于系统盘 二台用于数据盘	Rack Storage 8×2 一台用于系统盘 一台用于数据盘	Rack Storage 8×2 用于数据盘
	本地系统盘	4.2Gb×2×2 (RAID 1)逻辑系统盘 4.2Gb	4.2Gb×2×2 (RAID 1)逻辑系统盘 4.2Gb	2.1Gb×2×2 (RAID 1)逻辑系统盘 2.1Gb
	共享数据盘	9.1Gb×12 RAID 50 逻辑数据盘 91Gb	9.1Gb×8 RAID 50 逻辑数据盘 54.6Gb	4.2Gb×8 RAID 50 逻辑数据盘 25.2Gb

2. **双机双控容错系统软件** ServerGuard

(1)服务进程：

①双服务器采用 TCP/IP 网络协议和用户联播

②双机后台对于客户——服务器网络用户透明

·网络服务：双机后台对于用户一端，由软件 ServerGuard 提供一个逻辑的 IP Address，任一用户上网只需要这一地址；当后台有一台服务器出现故障时，另外一台服务器会自动将其网卡的 IP Address 替换为 ServerGuard 提供的逻辑地址，这样，用户一端的网络不会因为一台服务器出现故障而断掉。

·数据库服务：当有一台服务器出现故障时，另外一台服务器会自动接管数据库 engine；同时启动数据库和应用程序，使用户数据库可以继续操作。其流程图如图 2 所示。

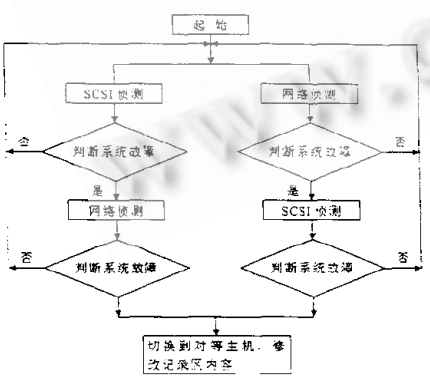

图 2

(2)监控原理：ServerGuard 内部含有 SCSI 侦测心跳及网络侦测心跳两条通信线路，结果置于 RS/8 磁盘柜上的一个 5MB 的小区。

·SCSI 监测：对于某一台服务器而言，将侦测信息以类似于记录方式写在该小区内，其中每一条记录包括如下内容：

①系统对本机的监测状态信息。

②另一台主机是否看到本机状态的信息。

当一台主机有问题或出现故障时，对等主机的可调度心跳频率不断提高，在最小心跳时间内发现记录内容没有更新、即会调用网络心跳侦测再次确认系统状态，当两组心跳都判断系统故障时，ServerGuard 将故障主机的交易业务在最小安全切换时间内切换到对等主机继续运行，同时修改记录区内容；一般切换时间不会超过 7 秒，根据应用程序的复杂程度，最小安全切换时间≤30 秒。

·网络侦测：业务主机对网络设备监测，同时配合 SCSI 心跳侦测，对等监控两台服务器主机的工作状态。当有一台服务器因为网络故障或其他原因引起故障而不能正常处理业务交易时，对等主机的可调度心跳频率不断提高；在最小心跳时间内发现心跳记录内容没有更新，即会调用 SCSI 心跳侦测再次确认系统状态；当两组心跳都判断系统故障时，ServerGuard 将故障主机的交易业务在最小安全切换时间内切换到对等主机继续运行，同时修改记录区内容；

五、结论

使用了双控系统后，满足了银行系统的高可用性、高安全性的要求，提高了使用银行的网络中数据的传输处理的能力，确实达到了银行提出的要求。

参考文献

[1] 《HP Net Server division》

[2] 《双机双控》，姜长安著

[3] 《双机双控容错系统》，Proware

[4] http:/www.hp/net server.com

[5] http:/www.proware/hot - standby.com

[6] http:/www.proware/Dataware.com

[7] 钟玉琢，《多媒体计算机技术》，清华大学出版社，1993

[8] T. Horignch，"An Optimization of modulation code in digital recording"，IEEE Vol. Mag - 12 No. 6 Nov. 1976

[9] 袁国兴，"大型计算机的信息输出和长期储存"，计算机数字通讯，1989. No. 6

(来稿时间：1998 年 12 月)