基于 MC/ServiceGuard 与 Oracle RAC 在首钢京唐公司能源管理系统中的应用研究

张冰锋
（北京首钢自动化信息技术有限公司电信事业部，唐山 063200）

摘　要　首钢京唐公司 EMS 系统成功运用 MC/ServiceGuard 集群实现了高可用性系统设计的基础上，针对 EMS 系统的特点、系统的硬件环境的设计、网络环境的设计、共享存储的设计、Oracle RAC 卷组的设计、MC/SG－Oracle 集群的设计和集群文件生成与配置等方面，介绍了 MC/ServiceGuard 集群系统的设计方法和系统文件配置。同时本文为 EMS 系统数据库稳定运行提供了可靠的技术保证。

关键词　EMS　MC/ServiceGuard　Oracle　集群配置文件

0　引言

北京首钢自动化信息技术有限公司利用国家大力发展钢铁厂能源管理系统（Energy Manage System，简称 EMS）建设的有利时机，同步开发实施首钢京唐钢铁联合有限责任公司（简称首钢京唐公司）能源管理系统。通过实施运营能源管理系统，完成公司内部能源数据共享，实现生产计划和能源调度的统一，达到了钢铁厂能源介质的生产控制/管理的信息自动化。能源管理系统的架构采用当代计算机和网络技术发展的成熟技术，运用多层信息化系统架构的设计理念，建立了分布式客户机/应用服务器/数据库服务器冗余协作处理环境的能源管理系统平台架构。采用开放的 UNIX 系统服务器作为系统的硬件配置，结合集群技术，实现应用服务器、Oracle 数据库服务器的并行处理、动态负载均衡和热备份的功能。达到了系统的高可用性和高可靠性的目标。

1　MC/ServiceGuard 概要

MC/ServiceGuard 是 HP－RX6600 动能服务器的高可用性集群，在计算机硬件和系统软件出现故障时，可继续执行应用程序的服务，是一种基于应用可迁移的方法。

多个通用系统通过网络相连，用 SCSI 或光纤通道总线连接需要物理共享的外存设备，但同时只有 1 台机器存取 1 个物理设备。系统间定时发送 heartbeat 信息，一旦对方系统故障，故障系统中的应用自动切换到备机上运行。

MC/ServiceGuard 管理应用是以 Package（应用包）为单位的，1 个 Package 包括 1 个浮动的 IP 地址、若干应用和系统进程及应用所用的硬盘。不管应用包如何切换，为前端用户提供服务所对应的 IP 地址是固定的，这样保证系统切换对用户是透明的。集群系统中没有 1 台机器空闲，各自运行自己的应用，使系统的能力得到了充分发挥。

MC/SeviceGuard 系统是系统硬件和集群软件相配合，完成单点故障的系统切换。集群的典型配置如图 1 所示。

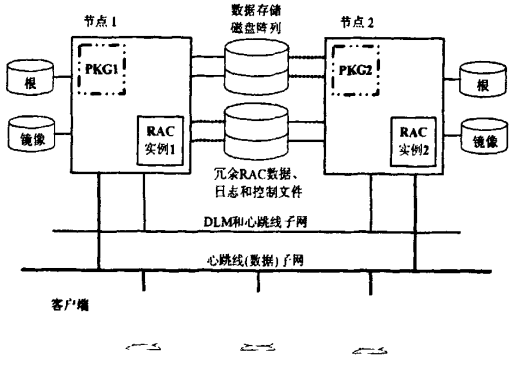

图 1　集群的典型硬件配置图

如果在集群系统中，节点 1 的硬件、服务、网络或其他资源出现故障，MC/SeviceGuard 则自动将程序包的控制权转移给集群中的另一节点 2，保证服务在系统中继续进行。系统的全部负荷由节点 2 承担。

集群切换后的硬件结构如图 2 所示。

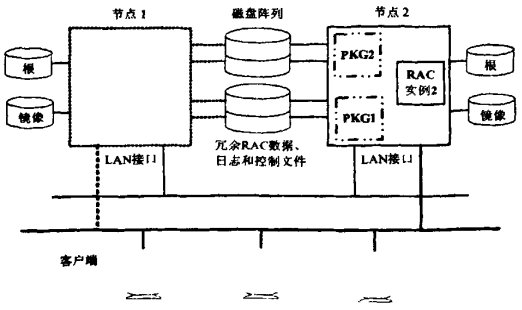

图2 集群切换后的硬件结构图

节点2将自动接受和控制节点1的程序包(PKG1)，程序包通常会一直留在节点2中。当节点1恢复正常后。可以通过手动和自动方式，将PKG1的控制权返回给节点1。

各个节点都有一组单独的磁盘与其相关联，其中含有程序包的应用程序以及所需的数据。为了防止磁盘组出现单点故障，一般都配置冗余磁盘提供数据的镜像副本。同时，共计有4个数据总线可供节点1和节点2相连的磁盘组使用。由于磁盘组都使用不同的总线，这样配置提供了最大冗余，还能提供最佳的I/O性能。

MC/ServiceGuard集群软件是在操作系统和磁盘卷组管理软件之上的系统软件，其层次结构如图3所示。MC/ServiceGuard组件是集群系统软件，程序包由用户根据不同的应用自行编制。MC/ServiceGuard共有9个守护进程，分别为配置守护进程、集群守护进程、日志守护进程、逻辑卷管理守护进程、Object Manager守护进程、SNMP子代理（可不运行）、服务助手守护进程、共享磁带守护进程和仲裁服务守护进程。

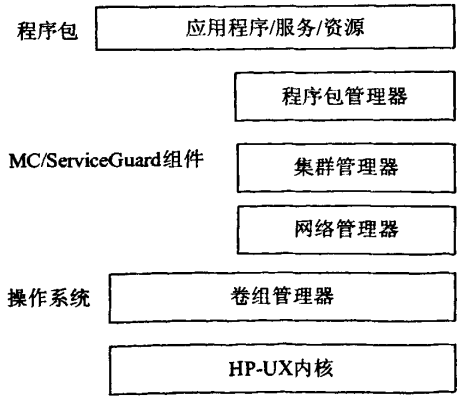

图3 集群软件层次结构图

2 MC/ServiceGuard在首钢京唐公司能源管理系统的实现方案

2.1 系统的特点

首钢京唐公司的能源管理系统是整合生产控制和生产管理的计算机系统，涉及过程监控、能源调度和能源管理等功能，形成能源生产控制和能源管理高度集成化的信息管理系统。能源生产控制和能源管理高度集成化，使数据库服务器需要支持来自AriEMS、SCADA、ERP、计量、MES等系统海量数据的频繁交易读写操作（见图4），联机交易数目超过3000。每个交易所涉及的关联数据表多，每个交易又需读取和修改大量的数据库表，所涉及的数据库表之间的关联相当复杂，且交易的实时性要求比较高，同时因为交易的结果直接影响EMS系统的可靠运行。所以处于联机交易数据核心的Oracle服务器的稳定性、可靠性必须得到保障。因此采用高可靠的数据库集群技术，提高了EMS系统的效率，通过负载均衡降低了网络和系统的单一开销，达到了应用数据稳定性。

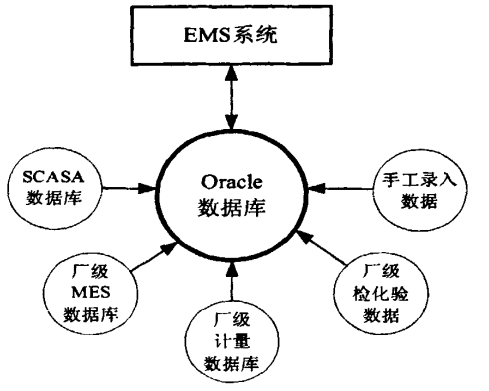

图4 能源管理系统数据结构图

2.2 硬件环境的构成

数据库服务器由2台可支持32路CPU的HP-rx6600主机组成，充分利用其优异的处理性能特点和硬件分区动态配置资源的特点，2台HP-rx6600并行工作并互为热备份。这2台HP-Rx6600之间通过MC/ServiceGuard组成高可用性集群系统，保证通信服务器的高可用性。集中存储系统由1台HP高端的EVA-4400全光纤交换式存储系统实现，HP Storage Works配置为1.4TB的可用存储容量并作镜像，16个光纤通道，8GB的镜像高速缓存。数据

备份系统采用 Tape Array 5300 磁带库，该磁带库有 4 个标准 LTO 驱动器（线性磁带驱动器），4 个磁带插槽，存储交换设备采用 2 台 16 口 HP SAN Switch 存储光纤交换机。这两组交换机之间实现动态热备和负荷均衡，并提供高达 32 口的连接能力。HP Storage Works 和 Tape Array 5300 磁带库均接在光纤交换机上，共享给交换机上层的所有主机设备，构成一个典型的存储局域网（SAN）环境，为今后系统发展和多平台集成提供了高性能、高扩展性的平台。

2.3 网络环境的设计

分别设计有 2 块千兆光纤网卡，与主干网的双网络交换机通过 1000Basesx 高速链路互连，交换机也通过热备份协议互连，从而保证数据中心网络无单点故障，各集群主机均另设计 1 块 100 兆网卡作为心跳网。通过 MC/ServiceGuard 的设计，互为备份的相对独立的两路千兆光纤网络在作为数据网络的同时，兼作心跳网络的备份，在集群主机间心跳网络出现故障时自动接管集群心跳网络服务，从而保证数据、心跳网络无单点故障，以此配置来保证了系统网络的高可用性。

2.4 共享存储的设计

Oracle RAC 集群中各主机作为主数据库服务器系统，都通过光纤交换机与 1 台 EVA4400 SAN 高端磁盘阵列相连，各主机采用共裸设备的方式激活 Volume Groups，通过 MC/ServiceGuard 与 Oracle10g RAC 集成组成集群系统，并行访问和存储共享数据。主机的本地磁盘设计为文件系统，作为 Oracle10g 数据库系统软件及数据库配置空间。

2.5 MC/SG - Oracle 方案的设计

由 HP - rx6600 的 2 个服务器分别组成节点 1 和节点 2，构成 MC/ServiceGuard 集群。运行 Oracle RAC 的 2 个并行的 RAC。MC/SG - Oracle 系统结构如图 5 所示。

Oracle RAC 集群中所有数据库的 VolumeGroups 的激活，由原来在程序包中启动，改成在数据库启动程序中启动，同样数据库服务启动前也通过程序作判断。如果 Oracle 数据库服务没有运行，则启动数据库服务，使 Oracle 数据库服务的开关闭不受程序包的控制。这样在系统维护时，集群可以停下来，数据库服务的还可以运行，同时在网络出现多点故障不能对外服务情况下，也不会出现宕数据库的现象，保证了网络恢复服务后，数据库可以立即恢复对外的服务。在 Oracle RAC 实现并行访问和共享磁盘的基础上，通过 MC/ServiceGuard 集群保证在系统上发生软件和硬件故障时，能将故障进行隔离并将服务切换到集群的其他服务器上。

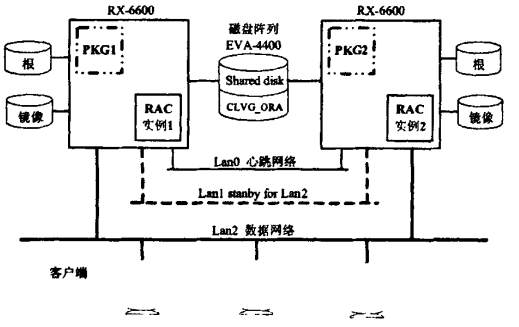

图 5 MC/SG - Oracle 系统结构图

2.6 集群文件生成和配置

2.6.1 集群配置文件

集群管理器管理的是一组定义集群参数的定义文件，这些参数保存在集群中所有节点上的二进制配置文件中，可以通过“cmquerycl”命令创建的集群配置模板文件，修改模板文件的相应参数，完成集群配置文件“cmclconf. ascii”。

#cmquery -v -C cmclconf. ascii -n clnode1 -n clnode2

2.6.2 集群程序包文件

集群的程序包文件，定义一组应用程序服务，当程序包在集群中的某个节点上启动时，由程序包管理器运行这组应用程序服务，该配置还按优先级顺序列出了可运行该程序包的集群节点，并定义了允许该程序包使用的可接受故障切换类型。程序包配置文件可以通过 SAM 的方式和“cmmakepkg”命令进行配置。

#cmmakepkg -p. pkgora1. ascii

2.6.3 集群程序包 控制文件

程序包必须有 1 个独立的、可执行控制脚本，放在程序包目录中，并且必须与程序包配置文件的 RUN_SCRIPT 和 HALT-SCRI 盯参数中指定的名称相同。可以用 cmmakepkg 命令进行生成模板文件，然后进行编辑。

#cmmakepkg, -scontrol. sh

2.7 集群启停与切换

在集群的各节点机上分别配置上述的集群配

置文件、程序包文件、程序包控制文件。

1）检查配置文件的正确性

#cmcheckconf -C/etc/cmcluster cmclconf. ascii

-P/etc/cmcluster/pkgora1/pkgora1. ascii

-P/etc/cmcluster/pkgora2/pkgora2. ascii

2）生成和分发二进制的文件

#cmapplyconf -v -C/etc/cmcluster/cmclconf. ascii

-P/etc/cmcluster/pkgora1/pkgora1. ascii

-P/etc/cmcluster/pkgora2/pkgora2. ascii

3）集群的启动。当完成配置文件检查、生成和分发后，使用 cmruncl 命令启动集群。其中可以通过 -n 选项来指定特定的节点。如果没有选项，将启动所有节点。

4）集群的关闭。集群的关闭，需要在集群所有节点上先停程序包，然后关闭所有集群节点。最后将整个集群关闭。

#cmhaltpkg pkgora1

#cmhaltnode clnode1

#cmhaltcl

5）查看集群状态信息

#cmviewcl

采用 cmviewcl 命令可查看集群的运行情况。在集群运行和关闭时，还可通过集群系统的 log 文件查看集群运行过程中，有无错误信息（在/etc/cmcluster/pkgora1 目录下，用#tail control. sh. log 查看）。

6）集群的故障切换。当集群中某个节点发生故障时，由于在集群的程序包文件定义了自动切换参数 AUTO_RUNYES，因此在发生故障时，程序包会自动切换到另一节点，程序包切换包括移动程序包和将它们的相关 IP 地址移动到新的系统中。故障切换时，原 TCP 连接将丢失，应用程序必须重新连接。当故障修复后，程序包的恢复有 2 种策略，AUTOMATIC 和 MANUAL（设置参数 FAILBACK-POLICY）。

故障的手动切换：程序包可以采用手动命令进行切换，用 cmhaltpkg 命令在一个节点停止程序包（#cmhaltpkg pkgora1）。用 cmiewcl 命令将看到 pkgora1 程序包的状态是 down，用 cmrunpkg 命令将程序包切换到另一个节点上。

3 结语

2008 年 12 月完成了在开发环境中搭建，MC/ServiceGuard 集群在首钢京唐公司能源管理系统中投用了近 1 年时间，按照系统的基本设计，实现防止单独故障和应用负荷均衡的目标。通过改变了 MC/ServiceGuard 集群的传统的设计模式，将所有数据库的磁盘卷组的激活由原来的在程序包中启动，改成在数据库启动程序中启动。程序包启动过程中，判断磁盘卷组是否被激活，如果没有则执行激活命令。通过这样的设计实现了集群停，而数据库不停的要求。

参考文献

[1] Oracle Real Application Clusters 10g [EB/OL]. 2006 - 01. http://www. oracle. com/ang/cn/technolo-gies/grid/index. html.

[2] 使用 ServiceGuard Extention For RAC. [EB/OL]. 2008 - 2. http://docs. hp. com/HighAvailability

[3] WILLIAM G, PAGE Jr. Oracle 开发使用手册 [M]. 李纪松，周保太等译. 北京：机械工业出版社，2000

[4] 王海军. ORACLE 学习教程 [M]. 北京：北京大学出版社，2000

[5] 马晓玉，孙岩等. ORACLE10g 数据库管理、应用与开发标准 [M]. 北京：清华大学出版社，2007

作者简介 张冰锋 助理工程师 从事能源管理软件的开发与实施工作