HP集群技术MC/ServiceGuard的应用

李世英　牛承佳　（中国联通广东分公司　510620）

摘要：本文介绍了HP集群技术MC/ServiceGuard在计费与客户服务系统中的应用及双结点集群系统的配置过程。对其他一些需要在高可靠性的环境下运行的关键进程，也可参考本文所介绍的方法进行处理。

关键词：集群　计费系统　心跳信号　包　服务

1　概述

随着电信市场的快速发展，竞争的日益激烈，不论是用户或者是电信运营商本身都对计费与客户服务系统(以下简称计费系统）不断提出更多、更高的要求。计费系统从总的方面来看应具备以下各种功能：话单采集、分拣、入库、传输、计费管理、结算管理、帐务管理、营业管理、销售管理、各类统计、备份管理、信用管理及停开机管理等等。在性能方面，计费系统应该具备非常高的可靠性和安全性，不论是任何哪一个方面出现问题，都会影响计费系统的可靠性，从而影响电信运营商对客户的服务质量，影响客户对电信运营商的忠诚程度，使电信运营商的声誉受到影响。

由于对计费系统可靠性和安全性要求的不断提高，如何建立起真正高可靠性的计费系统，目前已成为迫切需求。在这种情况下，集群作为一种提高性能及解决高可靠性要求的技术，得到了广泛的应用。

2　HP集群技术的特点

HP的MC/ServiceGuard企业集群为支持关键业务应用提供了一个强有力的支持。MC/ServiceGuard是专门用来保护关键业务应用免遭软、硬件故障的影响。在企业集群里，MC/ServiceGuard监视所有硬件和软件，检测故障，故障发生时进行快速响应，并为关键业务分配新的资源。检测故障并快速恢复应用的过程完全是自动的，不需操作员作任何干预。MC/ServiceGuard监视系统处理器、内存、LAN介质、LAN网卡、系统进程和应用程序进程，对故障快速响应，从而对基于LAN的客户恢复应用服务。

在企业集群中，MC/ServiceGuard不仅使应用程序可靠，而且采用特别方法保护数据完整性。当应用程序包从故障结点移出时，集群中其他结点互相协调确保失效结点不会危及应用数据的完整性。每个结点都知道集群中的其他成员及分配给它们的软件包。如果一个结点发生故障，剩下的结点会把其从集群中隔离出来以防止其访问磁盘。这一重要功能可以防止一个结点发生故障挂起或重启动后，不会再对现在已有别的结点负责的数据进行改写。

3　HP集群技术在计费与客户服务系统中的应用

在计费系统中，有很多应用进程都需要有高的可靠性，这里以其中的三个应用进程为例，叙述集群技术的实现过程。这三个应用进程是：/user1/home1/billing/comm：通信进程，负责对计费系统运行过程中产生的中间数据或结果数据在机器之间进行传送。

/user1/home1/billing/Sort：分拣进程，对原始话单进行分拣处理。

/user1/home1/billing/Input：入库进程，对分拣的结果入库。

3.1　集群的硬件连接

HP集群技术的硬件连接方案有很多种，本文涉及到的计费系统中，HP集群技术的硬件连接方案如图1所示，以两台HP N4000为主组成两个结点的集群系统作为计费系统的主处理系统，HP-UNIX版本为11.0，每台主机配置三块网卡，通过与CISCO SWITCH 5505相连构成局域网络，两台HP N4000主机均通过光纤通道与磁盘阵列相连，当其中一台主机出现问题时，另一台主机可以接管前者的进程。

HP N4000 (1)的三块网卡IP地址配置分别为:

A1: 192.1.1.1, 子网掩码: 255.255.255.0

B1: 空缺

C1: 193.2.1.1, 子网掩码: 255.255.255.0

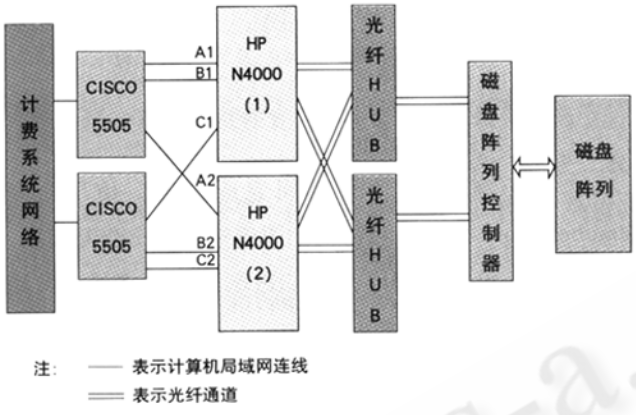

图 1 利用集群技术组成的计费系统
(两台N4000组成的集群系统)

HP N4000 (2)的三块网卡IP地址配置分别为:

A2: 192.1.1.2, 子网掩码: 255.255.255.0

B2: 空缺

C2: 193.2.1.2, 子网掩码: 255.255.255.0

通过CISCO 5505的配置,使得结点A1和结点A2属于LAN0,结点B1和结点B2属于LAN1,结点C1和结点C2属于LAN2,如图2所示。LAN0专门用来传送心跳信号,通过心跳信号,两台主机可以互相监视对方的运行状态,在必要时,接管对方的某些进程。守护进程cmcld在集群中的每个结点上运行,彼此监视对方的心跳信号,根据心跳信号判断对方是否还存活,当在一定时间后未收到从对方结点来的心跳信号,集群将把无心跳信号的结点删去。LAN2用来做数据传送,同时做为心跳信号的冗余备份传输通道。LAN1作为LAN 0和LAN2的后备,当LAN 0和LAN2任何一方出现故障时,可由LAN1接替。

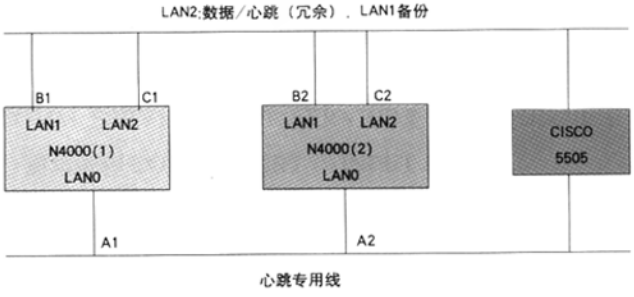

图 2 集群系统逻辑示意图

例如,假设在网络轮巡的过程中,cmcld进程检测到N4000(1)主机的LAN0发生故障,则cmcld将LAN0的IP地址转移到LAN1,此时,LAN1发出ARP广播,通知网络上的所有系统更新其本地的ARP缓冲区中IP—MAC地址对应关系,原LAN0的IP地址目前与LAN1的MAC地址相对应。广播完成之后,N4000(2)主机上的LAN0可继续与N4000(1)主机通信,N4000(1)主机端的IP地址不变,但此时的IP地址是LAN1的IP地址。为了完成上述动作,网络设备也需做一定的配置。

集群在形成或重组时,采用选举协议选出一个结点做为集群的协调员,负责处理集群中的各种事务。

3.2 集群的配置

3.2.1 MC/ServiceGuard中的逻辑卷管理(LVM)

计费系统中所使用的数据均存储在磁盘中,对磁盘的管理采用LVM方式,取代传统的磁盘分区方式,可以使系统管理员很方便地控制磁盘分区的大小和逻辑卷的大小。通常情况下,当一个集群中的两个结点均采用标准LVM同时激活一个卷组时,这两个结点可同时将文件系统挂上,于是,两个结点都可对共享文件系统进行读写操作,结果共享文件系统中的数据很快会被破坏。

为了解决该问题,MC/ServiceGuard对标准LVM做了修改。当一个卷组是一个MC/ServiceGuard集群的一部分时,在一个时间里只允许一个结点访问该卷组。

在LVM的头结构中,其中有一位在标准LVM中没有用到,在MC/ServiceGuard集群中,当这一位被置位时,表示卷组为MC/ServiceGuard的一部分,则在同一时间里只允许一个结点访问该卷组。

上述机制的实现是由守护进程cmlvmd来保证的。当集群激活时,该进程一直保持运行。当一个结点通过命令vgchange来激活卷组时,该命令首先被传到cmlvmd守护进程,该进程发一广播消息到所有其他结点上的cmlvmd守护进程,询问该卷组是否已被激活,若已有一结点激活该卷组,该结点的cmlvmd进程会给出相应的响应,结果是试图激活卷组的请求失败。当所有其他cmlvmd均响应该卷组没有被激活时,该卷组才可被激活。

在图1所示的计费系统中,MC/ServiceGuard中LVM的配置步骤如下,假设在n4000(1)主机上已生成一卷组/dev/vg01,卷组vg01上存储comm、sort和input三个应用进程需要调用的数据,在卷组上有一逻辑卷lv01挂在目录/user1下。

将vg01的VGID存到文件lvm_map中

vgexport _p _s _m/tmp/lvm_map /dev/vg01

· 将文件 lvm_map 传送到 n4000(2)主机，日的地路径除主机名之外，其他与源路径相同

rcp /tmp/lvm_map n4000(2):/tmp/lvm_map

· 在 n4000(2)主机上为卷组配置同 n4000(1)主机相同的路径

mkdir /dev/vg01

· 在 n4000(2)主机上为卷组 vg01 生成卷组文件，卷组参数与 n4000(1)主机上的卷组 vg01 相同

mknod/dev/vg01/group c 64 0x010000

· 在 n4000(2)主机上，重复在 n4000(1)主机上对卷组和逻辑卷所做的定义，以使两台主机可以对同一卷组及逻辑卷操作

vgimport _s-m /tmp/lvm_map /dev/vg01

· 在 n4000(2)主机上生成与 n4000(1)主机上相同的安装点

mkdir /user1

LVM配置完成之后，下一步配置集群。先配置无包集群，然后对包（Packages）和服务(Services)进行配置。

3.2.2 无包集群的配置

· 建立信任主机

在 HP UNIX 中，配置信任主机是通过超级用户的 .rhosts 文件，将所有需要信任的结点的名字（包括本地结点）写入该文件。在 MC/ServiceGuard 中，采用文件 /etc/cmcluster/cmclnodelist 替代超级用户的 .rhosts 文件，MC/ServiceGuard 在检查文件 .rhosts 之前先检查文件 cmclnodelist。这样做可以阻止集群内结点间的自由通信，把集群当作一个逻辑单元来保证其安全性。

· 生成由 n4000(1)主机与 n4000(2)主机组成的集群的 ASCII 配置文件 cmclconf.ascii，命令如下:

cmquerycl _C cmclconf.ascii _n N40000(1) _n N4000(2)

集群ASCII配置文件cmclconf.ascii须在/etc/cmcluster 路径下生成，该文件的内容包括很多项，主要如下:

* CLUSTER_NAME cluster1；集群名为cluster1。

* FIRST_CLUSTER_LOCK_VG /dev/gv01；锁盘所在的卷组为vg01。当集群由于故障原因，被分为含有相同结点数的两个子系统时，两个子系统争夺集群锁，获得集群锁的子系统成为新的集群，为了保证现有的应用可以在新的集群上正常运行，未争到集群锁的子系统被停下来。

* NODE_NAME N4000(1)及 N4000(2)；集群系统中的结点的主机名。

* HEARTBEAT_INTERVAL 1000000；结点向集群协调员发送心跳信号的时间间隔（微秒）。

* NODETIMEOUT 2000000；当超过该时间间隔（微秒）未收到某结点的心跳信号时，该结点将被认为无法再使用，集群启动一次重组。

* VOLUME_GROUP /dev/vg01；MC/ServiceGuard所使用的卷组，使用命令 vgchange _c y 将卷组标识为集群所使用的卷组。

· 编译并分发二进制文件

ASCII 配置文件编辑完成之后，使用命令 cmcheckconf 检查语法错误和逻辑错误，检查无误后，使用命令 cmapplyconf 产生二进制文件并将其分发到集群内其他结点。

3.2.3 集群中包（Packages）和服务(Services)的配置

在MC/ServiceGuard 高可用性环境中运行的应用必须同其所用到的资源一起配置在一个 Package 中。在 MC/ServiceGuard 集群中运行一个 Package 所需要的信息包含在包配置文件和包控制脚本中。每个Package都需要有一个包配置文件和包控制脚本。包配置文件中定义如子网或服务进程等应用所要用到的东西，以及Package的属性和特征。运行包控制脚本可以启动或停止一个包应用。

在图 1 所示的计费系统中，我们将三个应用进程 comm、sort 和 input 配置在同一个包 Package1 中，即包 Package1中包含有三个应用，该包使用pkg.conf和pkg.cntl 两个文件先在 N4000(1)主机上运行。由于卷组 VG01 中含有Package1 中的应用所需要的数据，所以，VG01 挂在 N4000(1)主机上。N4000(1)主机的网卡c1 有两个IP地址，一个是其本身的 IP 地址 193.2.1.1，子网掩码: 255.255.255.0，另一个是 Package1 的 IP 地址 193.2.1. 3，子网掩码: 255.255.255.0，如图 3 所示。

图 3 MC/ServiceGuard 集群中及包

当N4000(1)主机发生故障，在NODE_TIMETIMEOUT 时间内 N4000(2)主机收不到 N4000(1)主机的心跳信号，

MC/ServiceGuard认为N4000(1)主机上的应用失败，将应用从N4000(1)主机转移到N4000(2)主机，在N4000(2)主机运行Package1的包脚本，运行结果将VG01从原来挂在N4000(1)主机转移到挂在N4000(2)主机上，同时，将Package1的IP地址赋给N4000(2)主机上的网卡c2，在N4000(2)主机上运行Package1中的服务进程，如图3虚线箭头所示。

当N4000(1)主机恢复并重起后，集群恢复由两个结点组成。但是Package1能否自动返回N4000(1)主机，取决于所设定的恢复政策。恢复政策有两种：手动（缺省）和自动。手动时，必须将应用停下来，由系统管理员执行cmhaltpkg Package1和cmrunpkg Package1两条指令。自动时，Package1能自动返回N4000(1)主机。

包的配置过程如下:

· 与包相关的文件应放在路径/etc/cmcluster下，因此，包配置的第一步先生成一个子目录/etc/cmcluster/pkg1，然后进入该子目录。

· 用commakepkg生成包配置文件模版。

Commakepkg _p pkg.conf

· 编辑包配置文件pkg.conf，主要编辑的项目及内容如下:

* PACKAGE_NAME PACKAGE1; 包的名字。

* NODE_NAME N4000(1); 主结点名字。

* NODE_NAME N4000(2); 次结点名字。

* FAILBACK_POLICY MANUAL; 手动恢复政策。

* RUN_SCRIPT /etc/cmcluster/pkg1/pkg.cntl; 包运行脚本。

* RUN_SCRIPT_TIMEOUT NO_TIMEOUT

* HALT_SCRIPT /etc/cmcluster/ pkg1/pkg.cntl; 包停止脚本;

* HALT_SCRIPT_TIMEOUT NO_TIMEOUT

* SERVICE_NAME

COMM_SERVICE（SORT_SERVICE及INPUT_SERVICE);

MC/ServiceGuard中的服务名，服务名COMM_SERVICE、SORT_SERVICE及INPUT_SERVICE分别与进程COMM、SORT和INPUT相对应。

* SUBNET 193.2.1.0;

MC/ServiceGuard若访问不到该子网，则包失败。

· 编辑包脚本文件pkg.cntl，主要编辑的项目及内容如下:

* Path=/sbin:/usr/bin:/usr/sbin:/etc:/bin

当执行控制脚本时，PATH变量的值。

* VG [0] =vg01 由控制脚本激活的卷组名。

* LV [0] ="/dev/vg01/lv01";FS [0] ="/user1";

FS_MOUNT_OPT [0] =""

LV: 由控制脚本挂上的逻辑卷名; FS: 当挂逻辑卷时，挂点的文件系统名; FS_MOUNT_OPT: 当把逻辑卷挂到文件系统路径上时，安装选项列表。

* LV_UMOUNT_COUNT=1

当包关闭时，每个文件系统的尝试拆卸次数。

* IP [0] ="193.2.1.3"

SUBNET [0] ="193.2.1.0"

包的IP地址及子网

* SERVICENAME [0] ="COMM_SERVICE"

SERVICE_CMD [0] ="/user1/home1/billing/comm&"

SERVICE_RESTART [0] =""

SERVICENAME: 控制脚本中要启动的服务进程名，该名字与包配置文件中的服务名相对应; SERVICE_CMD: 用来启动服务进程的命令; SERVICE_RESTART: 当服务进程失败时，重启该进程的次数，空缺表示不重启。

在这里我们只配置了COMM_SERVICE，类似地，可以对SORT_SERVICE及INPUT_SERVICE进行配置。

* Function customer_defined_run_cmds

当启动包时先要运行的特定的应用进程命令。

* Function customer_defined_halt_cmds

当包停止后要运行的特定的应用进程命令。

· 将控制脚本及其他与包相关的文件拷贝分发到集群的另一个结点上，包文件在两个结点上的名字及路径要完全相同。

· 用命令cmcheckconf检查包文件的语法错误和逻辑错误，当检查没有错误时，用命令cmapplyconf生成并分发二进制文件。

· 用命令cmruncl启动集群。

4 小结

前面介绍了HP集群技术MC/ServiceGuard在计费系统中的应用及其配置过程，完成上述配置工作后，本文中所提到的三个在计费系统中很重要的进程comm(通信进程)、

（下转第59页）

(上接第55页)

Sort(分拣进程)和Input(入库进程)可实现在N4000(1)主机和N4000(2)主机之间的快速切换，当一台主机发生故障时，这些进程可在另一台主机上运行，从而保障关键处理的安全可靠。计费系统中除了本文中提到的三个关键进程之外，还有一些其他比较关键的进程需要在高可靠性的环境下运行，为了保证这些进程的可靠运行，也可参考本文所介绍的过程进行处理。

参考文献

1 *UNIX系统管理　卢显良　清华大学出版社　1993年8月*

2 *计算机网络原理与技术　丁正铨等　四川大学出版社　1995年9月*

3 *浅谈集群技术的应用　方华锋　计算机世界 20(c), 2000HP集群技术MC/ServiceGuard在计费与客户服务系统中的应用*