

AIXCHINA

HACMP for AIX 学习笔记

www.aixchina.com

AIX 中国论坛发表的所有文章版权均属相关权利人所有，受《中华人民共和国著作权法》及其它相关法律的保护。

如出于商业目的使用本资料或有牵涉版权的问题请速与论坛管理员联系。管理员电子邮件：aixchina@21cn.com

Aix HACMP

IBM Certification Study Guide Test 167

Chapter8 群集管理	5
8.1 监视群集	5
8.1.1 C1stat	5
8.1.2 HAView	5
8.1.3 log 文件	5
8.2 在节点或客户机启动和停止 HACMP	6
8.2.1 HACMP 守护进程	6
8.2.2 在节点启动 HA	6
8.2.3 在节点停止 HA	6
8.2.4 在客户机启动停止 HA 服务(clinfo)	7
8.3 更换故障部件	7
8.3.1 节点	7
8.3.2 卡	7
8.3.3 硬盘	7
8.4 共享 LVM 内容的修改	7
8.4.1 手工改	8
8.4.2 Lazy 修改	8
8.4.3 C-SPOC	8
8.4.4 Task Guide	9
8.5 修改群集资源组	9
8.5.1 加删改资源组	9
8.5.2 同步群集资源	9
8.5.3 DARE 资源迁移	9
8.6 安装 PTF	10
8.7 备份策略	10
8.7.1 Split-Mirror 备份	10
8.7.2 用事件安排定时备份	11
8.8 用户管理	11
8.8.1 列出所有群集节点用户	11
8.8.2 建新用户	11
8.8.3 改用户	11
8.8.4 删用户	11
8.8.5 组管理	12
8.8.6 C-SPOC log	12

AIX中国论坛

Chapter8 群集管理

8.1 监视群集

用/usr/sbin/cluster/clstat 检查群集状态，同时观察/tmp/hacmp.out，用 SMIT Show Cluster Services Screen 观察守护进程状态。

8.1.1 Clstat

Clstat 是 Clinfo 的客户程序，所以 Clinfo 要先在各节点运行。
/usr/sbin/cluster/etc/clhosts 要做好配置，包含各节点。
可运行于字符终端和 X-window，在 X 环境中，用-a 标志可出字符界面。

8.1.2 HAView

HAView 挂在 Netview 之中，要求 SNMP，各节点 SMUX peer daemon 要运行。

8.1.3 log 文件

8.1.3.1 /var/adm/cluster.log

错误、事件。

8.1.3.2 /tmp/hacmp.out

配置和起动脚本的输出是 cluster.log 的延伸，默认是 Debug Level run-time para high, verbose 状态。

8.1.3.3 /usr/sbin/cluster/history/cluster.mdd

细节的描述，用于错误处理。

8.1.3.4 System Error Log

8.1.3.5 /tmp/cm.log

clstrmgr 的活动，IBM 人员使用。

8.1.3.6 /tmp/cspoc.log

C-SPOC 命令的活动。

8.1.3.7 /tmp/emuhacmp.out

事件模拟的输出，也可用环境变量 EMUL OUTPUT 指定别的文件。

8.1.3.8 HACMP/ES 用到的三个 log

/var/ha/log/grpsvcs.<filename> grpsvcs 守护进程的活动

/var/ha/log/topsvcs.<filename> topsvcs 守护进程的活动

/var/ha/log/grpglsm grpglsm 守护进程的活动

8.2 在节点或客户机启动和停止 HACMP

8.2.1 HACMP 守护进程

- 8.2.1.1 clstrmgr 所有节点
管理心跳，监视节点，接口，执行相应的 scripts。
- 8.2.1.2 clsmuxpd 要求 snmpd，所有节点
维护群集的状态信息。
- 8.2.1.3 cllockd 可选
用于 Concurrent access 配置，但不是必要的。
- 8.2.1.4 clinfo 任意节点或客户机均可
是 clstat 所必需的，它会执行/usr/sbin/cluster/etc/clinfo.rc

With RSCT CRIS(system Cluster Technology)on HACMP/ES V4.3

- 8.2.1.5 Cluster Topology Services daemon (topsvcsd)
所有 HACMP/ES 节点要运行，监视网卡状态。
- 8.2.1.6 Cluster Event Management daemon (emsvcsd)
一个域中所有节点，监视系统资源。
- 8.2.1.7 Cluster Group Services daemon(grpsvcsd)
所有 HACMP/ES 节点要运行，管理所有群集操作要求的 distributed 协议。
- 8.2.1.8 Cluster Globalized Server Daemon daemon (grpqlsmd)
是 grpsvcsd 的客户进程，让 switch 卡可被各节点访问。

8.2.2 在节点启动 HA

用 SMIT Start cluster Services 建立和执行/usr/sbin/cluster/etc/rc.cluster 脚本，该脚本设置好环境变量，调用/usr/sbin/cluster/utilities/clstart 脚本，该脚本通过 startsrc 命令启各 HA 守护进程，子系统、组。

也可用 C-SPOC 工具，/usr/sbin/cluster/utilities/cl_rc.cluster。在某个节点或客户端串行启动各节点的 HA。（调用各节点 clstart cluster）

8.2.2.1 自动启 HA

通过 SMIT Start Cluster Services 或 rc.cluster -R 可在 inittab 中加入
hacmp:2:wait:/usr/sbin/cluster/etc/rc.cluster-boot > /dev/console 2 >&1

注意：不利于节点出问题后，先解决再加入 HA 的处理。

8.2.2.2 IPAT enable

rc.cluster 调用/etc/rc.net 做网络配置。

8.2.3 在节点停止 HA

通过 SMIT 建立和执行/usr/sbin/cluster/etc/clstop 脚本，该脚本通过 stopsrc 来停止 HA 的守护进程。

或用 C-SPOC 工具，cl_clstop 调用各节点 clstop 串行停止 HA。

8.2.3.1 何时停止

软硬件维护，重配置时。

8.2.3.2 停止类型

Graceful：HA 停止守护进程，释放资源，别的节点不接管。

Graceful with Takeover：同上，但别的节点接管资源。

Forced：只停守护进程，不释放资源：如 service IP，应用不停，fs 不会 unmount，sharevg 不会 varyoff，别的节点也不会来接管。HACMP/ES 不支持这种类型，只有传统 HA 有支持。

8.2.3.3 非正常守护进程中止

如果 SRC 检测到 daemon 异常中止，将用/usr/sbin/cluster/utilities/clexit.rc 停止系统，避免不可预期问题，不能用 kill-9 停止 clstrmgr，会引起系统停止。

8.2.4 在客户机启动停止 HA 服务(clinfo)

可用 rc.cluster 或 startsrc -s clinfo 启动

用 stopsrc -s clinfo 停止

8.2.4.1 clhosts 文件

客户端运行 clinfo，要先配好/usr/sbin/cluster/etc/clhosts 要包含所有节点的主机名或 IP(boot/service/standby)，Clinfo 将和这些节点的 clsmuxpd 通讯。

8.3 更换故障部件

8.3.1 节点

注意：内存足够，单 CPU 应用在 SMP 上可能更慢，槽位够吗，卡的位置，应用 lic 和 CPU ID 有关吗？

8.3.2 卡

先查线、hub 的问题，SSA 线可在线更换，SCSI 卡注意堵头和 ID。

8.3.3 硬盘

略

8.4 共享 LVM 内容的修改

建 VG，扩展、减少、改变、删除 VG，Importing，镜像，不镜像，同步 VG；

建 LV , 扩展、减少、改变、拷贝、删除 LV 或 LV 的拷贝 ;
建 FS , 扩展、改变、删除 FS ;
要在其它节点做 exportvg 和 importvg 的操作 , 来同步 VG DA 和 ODM 有 4 种途径

8.4.1 手工改

比如删 VG 必须手工在每个节点做(exportvg)。

步骤 :

停 HA (或和 SVG 有关的应用) , 改动 LVM
umount fs
vargoffvg
在另外节点 ,
exportvg
importvg from one of its disks , 注意 Vgmajor number
改 VG not auto varyon
mount fs
test fs
umount fs
varyoffvg
在其它节点重复 , 重启 HA

8.4.2 Lazy 修改

HA 可以自动完成 export 和 import VG 的工作 , 当 HA varyon 一个 VG 时 , 会比较 VGDA 中的时间片和节点/usr/sbin/cluster/etc/vg 文件中记录的时间片是否一致 , 如不一致 , HA 会在激活 VG 之前先做 export 和 import , 这会使接管动作多花几分钟 , 称之为 Lazy Update。

注意 :

(1)如 SVG 中第一个盘被换过 , Lazy Update 的 import 会出错 , 因为 ODM 中该盘的 PVID 和实际对不上 , importvg 是自动从第一个盘中取 VGDA。

(2)SCSI RAID 支持多个 LUN 的也有类似问题 , 因每个 LUN 对节点而言 , 就是一块盘。

8.4.3 C-SPOC

可以管理用户和组 , 维护 SLVM , 控制 HA。

可避免手工在各节点执行命令引起的人为问题。在一个节点执行的命令 , 同时也在其它节点执行 , 比如建一个用户 , 用 C-SPOC 命令可同时也在其它节点也加上。

C-SPOC 可通过 SMIT HA Cluste System Management 调用 , smit cl_lvm ; smit cl_conlvm 共同访问 vg。

8.4.4 Task Guide

8.4.4.1 要求

图形终端，已配好的 HA。

8.4.4.2 启动

命令行/usr/sbin/cluster/tguides/bin/cl_ccvg 或 smit hacmp cluster sys Management
cluster LVM Taskguide for Creating a Shared VG

8.5 修改群集资源组

资源为便于管理配置，被分成组，可控制资源组的节点又有优先级，这些定义的修改包括加、删、改资源组。

8.5.1 加删改资源组

不须要停止 HA，就可以做加删改的操作。

8.5.2 同步群集资源

smit 菜单 synchronize Cluster Resource

客户化的信息如 pre、post、notify、recovery 事件脚本的名字会被同步，因此各节点应有这些脚本，内容可以不同。

如本地节点 cluster manager 是不活动的，只 copy ODM 在 DCD(Default Configuration Dir)中的内容到远程节点。

如本地节点 clmanagr 是活动的，动态重配置事件开始，不仅 DCD、ACD (Active) 中的内容也被 Copy，且守护进程会刷新，在 log 文件中，有 reconfig resource-release、reconfig resource-acquire、reconfig resource-complete 三个事件被记录。

如本地节点 Clmanager 未激活，但其它节点有激活，同步失败。

8.5.3 DARE 资源迁移

Dynamic Reconfiguration Resource Migration utility 可以方便地控制一个资源组从一个节点到另一个节点，或禁止要求一个资源组，而不影响节点，HA 的其它正常活动。

8.5.3.1 资源迁移的类型

Sticky 资源迁移，让一个资源组习惯性地属于某一节点，除非节点 Down，高优先级节点加入，资源组不会属于该节点。一般用于 cascading 类配置，不用于 rotating 配置。

Non-Sticky 资源迁移：让资源组临时性属于某一节点，适用于 rotating 配置，或 Cascading Inactive Takeover=false 时用。

8.5.3.2 Locations 位置

可以用节点名填入 Location 域，说明 Sticky 和 Non-Sticky 的归属节点。也可用关键字 Default 和 stop(DARE 工具支持)来填 Location。

Default：清除以前的 stickiness 设置，让资源组按默认的 Cascading 和 rotating 行为习惯活动。如空着 Location 域，就是 default。

当一个 Cascading 资源组，INACTIVE_TAKEOVER false，而主节点没开机，该资源组不会被激活，当用 Default 迁移可让活动的最高优先级的节点拿到该资源组。

Stop：让一个资源组 stop，且不会被任何失败和重加入事件导致活动(要带 sticky)。

8.5.3.3 cldare 命令

cldare：同时执行所有的迁移请求，有一个不行，则所有迁移都不会发生。

cldare -m <resgroup name>：[Location |default stop nodename][sticky]

-M 代表迁移，nodename 是 up 的且在 resgroup 的 nodelist 中，可重复多个资源组。

一般来讲 stop 和 sticky 同用，避免 reboot 时启动资源。

但 stop 和 non-sticky 用于停下 INACTIVE-TAKEOVER false 的 cascading 资源组。

8.5.3.4 clfindres

可查找一个资源组的状态。

8.5.3.5 清除 sticky 标志(记在 HA OIM 中)

当 HA 是 UP 时，可用 Default 关键字，或 non-sticky 迁移

当 HA 是 Down 时，用 cldare -v(不校验) -M <resgrp name>：stop

8.6 安装 PTF

apply AIX PTF 可以让 B 机 Takeover A 机应用，先装 A，再将 B，关键应用不会停止。

- 1、smit clstop with takeover
- 2、Apply PTF
- 3、/usr/sbin/cluster/diag/clverify 和其它 Test
- 4、reboot node，如果涉及 cluster.base.client.lib 而应用到 Cluster Lock Manager 或 Clinfo API function，要重新 relink 应用。
- 5、Restart HA:smit clstart 接管回资源
- 6、在 B 机重复 1-5

注意：(1)群集节点要求保持 AIX maintenance Level 一致，但 update 时例外；

(2)群集节点要求 HA maintenance level 一致，且 update 时要求 Down 掉 HA

8.7 备份策略

8.7.1 Split-Mirror 备份

由于备份要求数据保持 offline，不能应用一边要修改，一边做备份，而有些应用又不允许长时间 offline 做备份，可用下法：

- 1、lsvg -l VGNAME 找到 lv 的名字和 fs 相关
- 2、Stop 应用 umount fs 注意:至少有一次镜像的 lv 才行
- 3、splitlvcopy -y newlv fslv 将原 lv 分成 2 个 lv
- 4、mount fs start 应用

- 5、用 newlv 建 fs(如何建?在/etc/file systems 中加一个 fs newlv),并 mount it read-only
- 6、备份 newlv 的 fs
- 7、删除 newlv
- 8、用 mk lvcopy 加 fslv 一个镜像,并同步恢复原状。

8.7.2 用事件安排定时备份

利用 crontab 做备份,当发生接管时,利用 node_down_remote 事件的 post-event 脚本修改 crontab,将接管资源的备份也包含进去,用 node_up_remote 事件脚本改回 crontab,备份脚本支持 split 备份。

8.8 用户管理

用 C-spoc 创建的用户可保证有相同的 uid 和 gid(在整个群集)

8.8.1 列出所有群集节点用户

- (1)lsuser (RSH)在每个节点执行;
- (2)cl_lsuser(C-SPOC)会自动执行 lsuser 在每个节点,如用了 NIS,有些 user 信息不会显示;C-SPOC 也有 SMIT 入口

8.8.2 建新用户

- (1)mkuser 相关文件 /usr/lib/security/mkuser.default
- (2)cl-mkuser

8.8.3 改用户

- (1)chuser
- (2)cl_chuser 当采用 NIS 时,不要用 cl_chuser

8.8.4 删用户

用户的帐号删除了,但用户的 home 目录保留,只有 root 和同组用户可访问。

- (1)rmuser
- (2)cl_rmuser

8.8.5 组管理

和用户管理类似。

8.8.6 C-SPOC log

/tmp/cspoc.log 只记录在本机运行 cspoc 命令情况。