

Hacmp 5.1 配置手册

一， Hacmp 的软件安装。

1. 安装前提

如果您的操作系统是 AIX 5 . 1 , 那么您要安装维护补丁包 M L 0 3 以上, 目前最高的补丁版本是 M L 0 5 . 并且您还要安装 RSCT 2.2.1.30 或更高版本。

以下的包也是必须要安装的:

```
bos.adt.lib
bos.adt.libm
bos.adt.syscalls
bos.net.tcp.client
bos.net.tcp.server
bos.rte.SRC
bos.rte.libc
bos.rte.libcfg
bos.rte.libcur
bos.rte.libpthreads
bos.rte.odm
```

如果您要安装并行的资源组, 还要安装下面的包:

```
bos.rte.lvm.rte5.1.0.25 or higher
bos.clvm.enh.
```

2. 开始安装

一般基本上除了 haviw , netwiw (Tivoli), 的包以外, 所有的 hacmp 的包都要安装。

3. 打补丁。

注意, 客户总是忽略给 hacmp 打补丁这一步骤。其实对 hacmp 来说, 补丁是十分重要的。很多发现的缺陷都已经在补丁中被解决了。有的客户严格的按照正确步骤安装和配置完 hacmp 的软件后, 发现 takeover 有问题, ip 接管有问题, 机器自动宕机等千奇百怪的问题, 其实都与补丁

有关。所以客户一定要注意打补丁这个环节。现在 hacmp 最新的补丁是:

IY53044 - Latest HACMP for AIX R510 Fixes as of January 2004

大家可以从 IBM 网站上下载。

4. 重启机器。在 hacmp 5.1 中 为了安全起见, 不再使用 /.rhosts 文件来控制两台机器之间的命令和数据交换, 而是引进的一个新的进程 clcmd 。 如果你编辑 /etc/inittab 文件

就会发现安装完 hacmp 后, 在最后添加了一行: clcmdES:2:once:starts -s clcmdES
>/dev/console 2>&1 。因此重新启机后, ps -ef |grep clcmd , 会发现 :root 12908 6478
0 Apr 12 - 0:21 /usr/es/sbin/cluster/clcmd -d , 证明该进程启动了。

Hacmp5.1 使用

/usr/es/sbin/cluster/etc/rhosts 文件来代替 /.rhosts 文件的功能。

注意: 如果两个节点间的通讯发生了什么问题, 可以检查 rhots 文件, 或者编辑 rhosts 文

件

加入两个节点的网络信息。

二, hacmp5.1 的配置

我们以两台机器为例: test1 和 test2, 共享三块 7133 硬盘。

1. 首先配置两台机器的 ip 和 vg, 以及 /etc/hosts 和 application 启动/停止脚本

test1: />netstat -in

Name	Mtu	Network	Address	Ipkts	Ierrs	Opkts	Oerrs	Coll
en0	1500	link#2	0.4.ac.49.f2.d5	77960	0	47805	0	0
en0	1500	100.1	100.1.0.1	77960	0	47805	0	0
en1	1500	link#3	0.6.29.ec.44.d6	33	0	11	0	0
en1	1500	192.168.0	192.168.0.1	33	0	11	0	0

test2: />netstat -in

Name	Mtu	Network	Address	Ipkts	Ierrs	Opkts	Oerrs	Coll
en0	1500	link#2	0.4.ac.49.60.23	31138	0	82582	0	0
en0	1500	100.1	100.1.0.2	31138	0	82582	0	0
en1	1500	link#3	0.4.ac.3e.b9.4b	36	0	13	0	0
en1	1500	192.168.0	192.168.0.2	36	0	13	0	0

test1 :

hdi sk0	0004383268b07574	rootvg	active
hdi sk3	000438325e22bca7	test1vg	
hdi sk4	00043832125e5aa8	None	
hdi sk5	000438323d0e4487	None	

test2 :

hdi sk0	000d29574085126d	rootvg	active
hdi sk5	000438325e22bca7	test1vg	
hdi sk6	00043832125e5aa8	None	
hdi sk7	000438323d0e4487	None	

/etc/hosts

100.1.0.2	test2_boot1	test2
100.1.0.1	test1_boot1	test1
192.168.0.1	test1_boot2	
192.168.0.2	test2_boot2	
10.1.0.1	test1_svc	
10.1.0.2	test2_svc	
10.1.0.5	test1_per	
10.1.0.6	test1_per	

test2: /ha51>ls -l

-rwxr-xr-x	1	root	system	65 Apr 13 13:51	start
-rw-r--r--	1	root	system	31 Apr 13 11:49	start.log

```

-rwxr-xr-x  1 root    system    66 Apr 13 14:01 start1
-rw-r--r--  1 root    system    31 Apr 13 14:01 start1.log
-rwxrwxrwx  1 root    system    64 Apr 13 11:48 stop
-rw-r--r--  1 root    system    31 Apr 13 11:48 stop.log
-rwxr-xr-x  1 root    system    66 Apr 13 14:01 stop1
-rw-r--r--  1 root    system    31 Apr 13 14:01 stop1.log

```

vi start

date >> /ha51/start.log

banner " start app1 " >> /tmp/hacmp.out

vi stop

date >> /ha51/stop.log

banner "stop app1 " >> /tmp/hacmp.out

vi start1

date >> /ha51/start1.log

banner " start app2 " >> /tmp/hacmp.out

vi stop1

date >> /ha51/stop1.log

banner "stop app2 " >> /tmp/hacmp.out

注意：在两个节点要保证 hosts 和 启动/停止脚本要一样存在。

2. 用 smitty hacmp 来配置 hacmp

添加 cluster 和 node

smitty hacmp

Initialization and Standard Configuration

Extended Configuration

System Management (C-SPOC)

Problem Determination Tools

Add Nodes to an HACMP Cluster

Configure Resources to Make Highly Available

Configure HACMP Resource Groups

Verify and Synchronize HACMP Configuration

Display HACMP Configuration

* Cluster Name [ha51tsc]

New Nodes (via selected communication paths) [m [test2_boot1
test1_boot1]

Currently Configured Node(s)

这一部很重要，一般我们都是把每个节点的 boot1 作为 communication path。New node 可以一起加，也可以一个一个的加。当回车以后，系统会自己 discover hacmp 的资源显示如下：

```

.....
IP Network Discovery completed normally
Current cluster configuration:

No resource groups defined
Cluster Description of Cluster: ha51tsc
Cluster Security Level: Standard
There are 2 node(s) and 1 network(s) defined

```

NODE test1:

```

    Network net_ether_02
           test1_boot1    100.1.0.1
           test1_boot2    192.168.0.1

```

NODE test2:

```

    Network net_ether_02
           test2_boot1    100.1.0.2
           test2_boot2    192.168.0.2

```

.....

添加高可用资源 (service ip , application server , vg and jfs)

添加服务 ip 地址

Add Nodes to an HACMP Cluster

Configure Resources to Make Highly Available

Configure HACMP Resource Groups

Verify and Synchronize HACMP Configuration

Display HACMP Configuration

Configure Service IP Labels/Addresses

Configure Application Servers

Configure Volume Groups, Logical Volumes and Filesystems

Configure Concurrent Volume Groups and Logical Volumes

Add a Service IP Label/Address

Change/Show a Service IP Label/Address

Remove Service IP Label(s)/Address(es)

* IP Label /Address	[test1_svc]
Network Name	[net_ether_02]
* IP Label /Address	[test2_svc]
Network Name	[net_ether_02]

添加 application server

Configure Service IP Labels/Addresses

Configure Application Servers

Configure Volume Groups, Logical Volumes and Filesystems

Configure Concurrent Volume Groups and Logical Volumes

Add an Application Server

Change/Show an Application Server

Remove an Application Server

* Server Name	[app1]
* Start Script	[/ha51/start]
* Stop Script	[/ha51/stop]
* Server Name	[app2]
* Start Script	[/ha51/start1]
* Stop Script	[/ha51/stop1]

添加共享 vg , jfs

注意在前面的步骤中我们看到已经有一个共享 VG test1vg 存在了 ,它使用传统的方法 :

1 . 在 test1 节点上创建 test1vg , lv,jfs

2 . Varyoffvg

3 . 在 test2 上 importvg

4 . Varyoffvg

现在我们试着用 hacmp 的功能去创建 test2vg

Configure Service IP Labels/Addresses

Configure Application Servers

Configure Volume Groups, Logical Volumes and Filesystems

Configure Concurrent Volume Groups and Logical Volumes

Shared Volume Groups

Shared Logical Volumes

Shared File Systems

Synchronize Shared LVM Mirrors
Synchronize a Shared Volume Group Definition

List All Shared Volume Groups
Create a Shared Volume Group
Create a Shared Volume Group with Data Path Devices
Set Characteristics of a Shared Volume Group
Import a Shared Volume Group
Mirror a Shared Volume Group
Unmirror a Shared Volume Group

在选择菜单中同时用 F7 选择 test1 和 test2

➤ test1

➤ test2

选中 PVID 00043832125e5aa8

Node Names	test1, test2
PVID	00043832125e5aa8
VOLUME GROUP name	[test2vg]
Physical partition SIZE in megabytes	4
Volume group MAJOR NUMBER	[49]

test2: /ha51>ls pv

hdi sk0	000d29574085126d	rootvg	active
hdi sk5	000438325e22bca7	test1vg	
hdi sk6	00043832125e5aa8	test2vg	
hdi sk7	000438323d0e4487	None	

test1: /ha51>ls pv

hdi sk0	0004383268b07574	rootvg	active
hdi sk3	000438325e22bca7	test1vg	
hdi sk4	00043832125e5aa8	test2vg	
hdi sk5	000438323d0e4487	None	

同样方法你可以在两个节点上同时创建 l j fs

Shared Volume Groups
Shared Logical Volumes
Shared File Systems
Synchronize Shared LVM Mirrors
Synchronize a Shared Volume Group Definition

Journalled File Systems
Enhanced Journalled File Systems

Add a Journal ed File System
 Add a Journal ed File System on a Previously Defined Logical Volume
 List All Shared File Systems
 Change / Show Characteristics of a Shared File System
 Remove a Shared File System

Add a Standard Journal ed File System
 Add a Compressed Journal ed File System
 Add a Large File Enabled Journal ed File System

test1vg	test1, test2
test2vg	test1, test2
Node Names	test1, test2
Volume group name	test1vg
* SIZE of file system	[10]
* MOUNT POINT	[/test1j fs]
PERMISSIONS	read/write
Mount OPTIONS	[]
Start Disk Accounting?	no
Fragment Size (bytes)	4096
Number of bytes per inode	4096
Allocation Group Size (MBytes)	8

系统会自动在 test1 上添加 test1j fs 文件系统，并且自动会在两个节点上作 update 。但是根据经验，最好还是用传统的方式在一个结点上创建 vg ，lv, j fs 。然后再 import 到另一个节点上。这里有一个 tips ，如果在这里创建共享 j fs 遇到问题，可以先手工把 vg 在一个结点上 varyon ，然后再创建就可以了。

创建资源组

Initialization and Standard Configuration
 Extended Configuration
 System Management (C-SPOC)
 Problem Determination Tools

Add Nodes to an HACMP Cluster
 Configure Resources to Make Highly Available
 Configure HACMP Resource Groups
 Verify and Synchronize HACMP Configuration
 Display HACMP Configuration

Add a Resource Group
 Change/Show a Resource Group
 Remove a Resource Group
 Change/Show Resources for a Resource Group (standard)

Cascading
 Rotating
 Concurrent
 Custom

* Resource Group Name [res1]
 * Participating Node Names / Default Node Priority [test1 test2]

同样方法可以添加 res2

接下来可以配置资源组,当然也可以在 Extended Configuration 中去详细配置。
 我们姑且先在 Initialization and Standard Configuration 中配置。

Smitty cm_config_hacmp_resource_groups_menu_dmn

Add a Resource Group
 Change/Show a Resource Group
 Remove a Resource Group
 Change/Show Resources for a Resource Group (standard)

选择 res1

Resource Group Name res1
 Participating Node Names (Default Node Priority) test1 test2
 * Service IP Labels/Addresses [test1_svc]
 Volume Groups [mtest1vg]
 Filesystems (empty is ALL for VGs specified)
 [/test1jfs]
 Application Servers
 [mapp1]

同样的方法配置 res2

检查和同步 hacmp 配置

Initialization and Standard Configuration
 Extended Configuration
 System Management (C-SPOC)
 Problem Determination Tools

Add Nodes to an HACMP Cluster
 Configure Resources to Make Highly Available
 Configure HACMP Resource Groups
 Verify and Synchronize HACMP Configuration
 Display HACMP Configuration

Cluster Description of Cluster: ha51tsc
 Cluster Security Level: Standard
 There are 2 node(s) and 1 network(s) defined

NODE test1:

Network net_ether_02	
test2_svc	10.1.0.2
test1_svc	10.1.0.1
test1_boot2	192.168.0.1
test1_boot1	100.1.0.1

NODE test2:

Network net_ether_02	
test2_svc	10.1.0.2
test1_svc	10.1.0.1
test2_boot1	100.1.0.2
test2_boot2	192.168.0.2

Resource Group res1

Behavior	cascading
Participating Nodes	test1 test2
Service IP Label	test1_svc

Resource Group res2

Behavior	cascading
Participating Nodes	test2 test1
Service IP Label	test2_svc

注意 nodetest1 的 ip 地址排列，虽然 test_boot2 排在 test_boot1 前面，但是实验证明，service 地址依然会绑定在 communication path 上。

现在就可以做 Initialization and Standard Configuration Verify and Synchronize HACMP Configuration .
然后 start 一下 hacmp ，看看 take over 是否都正常。

注意，很多客户是把所有的 hacmp 包括应用都配好后再试起 hacmp ，作 takeover 测试，这是很不好的一种习惯。因为融入的可能因素太多了，一旦有了问题，我们还要隔离问题，先把 hacmp 配置简化，再一步步作 pd ，那么先前的配置就白做了。所以建议客户阶段性的监测一下 hacmp

3 . 到此为止我们的 hacmp 已经基本配置完成了。剩下的要在 Extended Configuration 中配置了。

在 Extended Configuration 中我们还可以配置 tty 心跳网络 ， hdisk 心跳网络，Persistent Node IP ，application monitor 等等。

下面我们先介绍一下配置 hdisk 心跳网络，这也是 hacmp5.1 里的一个新的功能。

首先我们要一个 Enhanced concurrent VG ，这个vg 不需要一定是放在 concurrent 资源组里的vg ，当然也可以用concurrent 资源组里的硬盘来做心跳网络。这个concurrent vg 可以通过传统方法建立。

1 . Mkvvg -c convg

2 . 在一个节点上varyoffvg ，另一个节点上importvg

现在我们介绍用hacmp 来创建concurrent vg .

Initialization and Standard Configuration
Extended Configuration
System Management (C-SPOC)
Problem Determination Tools

Add Nodes to an HACMP Cluster
Configure Resources to Make Highly Available
Configure HACMP Resource Groups
Verify and Synchronize HACMP Configuration
Display HACMP Configuration

Configure Service IP Labels/Addresses

Configure Application Servers
 Configure Volume Groups, Logical Volumes and Filesystems
 Configure Concurrent Volume Groups and Logical Volumes

Concurrent Volume Groups
 Concurrent Logical Volumes
 Synchronize Concurrent LVM Mirrors

List All Concurrent Volume Groups
 Create a Concurrent Volume Group
 Create a Concurrent Volume Group with Data Path Devices
 Set Characteristics of a Concurrent Volume Group
 Import a Concurrent Volume Group
 Mirror a Concurrent Volume Group
 Unmirror a Concurrent Volume Group

选中 test1 and test2

选中共享硬盘

Node	Names
test1, test2	
PVID	
000438323d0e4487	
VOLUME GROUP name	convg
Physical partition size in megabytes	4
Volume group MAJOR NUMBER	[49]
Enhanced Concurrent	Mode
true	

下面看一下两个节点的硬盘状况：

test1:			
hdisk0	0004383268b07574	rootvg	active
hdisk3	000438325e22bca7	test1vg	
hdisk4	00043832125e5aa8	test2vg	
hdisk5	000438323d0e4487	convg	
test2 :			
hdisk0	000d29574085126d	rootvg	
active			
hdisk5	000438325e22bca7	test1vg	
hdisk6	00043832125e5aa8	test2vg	
hdisk7	000438323d0e4487	convg	

现在检查 hdi sk 网络的状况，在一个节点上向 hdi sk 写数据，从另一个节点上读数据，很像 我们在配置 tty 网络之前，检查一下 tty 是否连通。

注意：我原来的操作系统是 aix5.2 01，安装了 hacmp5.1 打了最新的补丁。但是在/usr/sbin/rsct/bin 下找不到 dhb_read 命令。它应该是属于 rsct 的，后来我把 aix5.2 打倒 ml02，rsct 所有的包都生级了，reboot 机器后，找到了 dhb_read 命令。

1. Add /usr/sbin/rsct/bin/ to /etc/environment 里的 path 中
2. 重新 login test1 和 test2 使 path 生效
3. 在 test1 上运行：dhb_read -p hdi sk5 -r
4. 在 test2 上运行：dhb_read -p hdi sk7 -t

在 test1 上：test1:/>dhb_read -p hdi sk5 -r

Receive Mode:

Waiting for response . . .

Link operating normally

在 test2 上：

test2:/usr/sbin/rsct/bin>dhb_read -p hdi sk7 -t

Transmit Mode:

Detected remote utility in receive mode. Waiting for response . . .

Link operating normally

证明通讯正常。

添加 hdi sk heart beat 网络和设备

Initialization and Standard Configuration

Extended Configuration

System Management (C-SPOC)

Problem Determination Tools

Discover HACMP-related Information from Configured Nodes

Extended Topology Configuration

Extended Resource Configuration

Extended Event Configuration

Extended Performance Tuning Parameters Configuration

Security and Users Configuration

Snapshot Configuration

Extended Verification and Synchronization

Configure an HACMP Cluster

Configure HACMP Nodes

Configure HACMP Sites
 Configure HACMP Networks
 Configure HACMP Communication Interfaces/Devices
 Configure HACMP Persistent Node IP Label/Addresses
 Configure HACMP Global Networks
 Configure HACMP Network Modules
 Configure Topology Services and Group Services
 Show HACMP Topology

Add a Network to the HACMP Cluster
 Change/Show a Network in the HACMP Cluster
 Remove a Network from the HACMP Cluster

Pre-defined Serial Device Types

diskhb
 rs232
 tm SCSI
 tm SSA

* Network Name [m [net_diskhb_01]
 * Network Type diskhb

添加设备：

Extended Configuration Extended Topology Configuration Configure HACMP
 Communication Interfaces/Devices Add Communication Interfaces/Devices Add
 Pre-defined Communication Interfaces and Devices Communication Devices
 net_diskhb_01

* Device Name [heartbeatdisk5]
 * Network Type diskhb
 * Network Name net_diskhb_01
 * Device Path [/dev/hdisk5]
 * Node Name [test1]

* Device Name [heartbeatdisk7]
 * Network Type diskhb
 * Network Name net_diskhb_01
 * Device Path [/dev/hdisk7]
 * Node Name [test2]

Extended Configuration Extended Topology Configuration Show HACMP Topology
 Cluster Description of Cluster: ha51tsc
 Cluster Security Level: Standard

NODE test1:

```

Network net_diskhb_01
    heartbeatdisk5 /dev/hdisk5
Network net_ether_02
    test1_svc      10.1.0.1
    test2_svc      10.1.0.2
    test1_boot2    192.168.0.1
    test1_boot1    100.1.0.1

```

NODE test2:

```

Network net_diskhb_01
    heartbeatdisk7 /dev/hdisk7
Network net_ether_02
    test1_svc      10.1.0.1
    test2_svc      10.1.0.2
    test2_boot1    100.1.0.2
    test2_boot2    192.168.0.2

```

配置永久的IP标识 (persistent IP label)

一个永久的IP标识 (persistent IP label) 是一个IP别名，它可以被分配给一个群集网络中的指定的节点，并且会一直固定在分配的节点上。

2. 永久的IP标识 (persistent IP label) 的特性：

- (1) 一直固定在被分配的节点上 (节点绑定)
- (2) 作为别名被配置在启动网卡 (boot adapter) 上
- (3) 与已经被配置的服务IP标识 (service IP label) 或启动IP标识 (boot IP label)

共同存在

- (4) 不需要在节点上安装额外的物理网卡
- (5) 不属于任何资源组
- (6) 可以被用于在群集中访问指定的节点进行管理工作
- (7) 在节点启动后即可用，当HACMP服务停止后也始终保持可用
- (8) 在以太网、令牌环网、FDDI以及ATM LANE网络中都可被配置
- (9) 不能在SP交换机、ATM传统IP网和串行网络上进行配置
- (10) 和配置的服务IP标识 (service IP label) 和启动IP标识 (boot IP label) 使用同一块网卡
- (11) 如果节点失败，该IP标识不会迁移到群集中的其它节点
- (12) 如果网卡失败，它只会迁移到相同网络的同一个节点上的其它网卡
- (13) 每个网络的每个节点上只能配置一个永久的IP标识 (persistent IP label)

3. 子网的要求

- (1) 对于使用传统的IPAT的网络 (不使用别名)
 - a. 必须被配置为和网络中该节点上的所有standby IP标识在不同的子网
 - b. 可以被配置为和网络中该节点上的service IP标识和boot IP标识在相同的子网或者是不同的子网

(2) 对于使用别名的IPAT的网络

- a. 必须被配置为和网络中该节点上的所有boot IP标识在不同的子网
- b. 可以被配置为和网络中该节点上的作为boot网卡别名的service IP标识在相同的子网或者是不同的子网

Extended Configuration	Extended Topology Configuration	Configure HACMP
Persistent Node IP Label/Addresses	Add a Persistent Node IP Label/Address	
* Node Name	test1	
* Network Name	net_ether_02	
Node IP Label/Address	test1_per	
* Node Name	test2	
* Network Name	net_ether_02	
Node IP Label/Address	test2_per	

注意：永久ip 同步完后，ip 立即绑定到boot1 上。

同步：Extended Configuration Extended Verification and Synchronization

启动 hacmp .

三 . Hacmp 的监控和问题诊断

1 . Clstat 监控 hacmp

首先加路径：/usr/es/sbin/cluster 到/etc/environment 的 path 中。

在 aix5.2 下要对 snmp 做一些调整才可以看到真正的 hacmp 的状态。

具体来说，aix 5.2 的 snmp 默认是 version 3：

```
test2:/usr/sbin>ls -l |grep snmp
lrwxrwxrwx  1 root    system      8 Apr 08 17:55 clsnmp -> clsnmpne
-rwxr-x---  1 root    system      83150 Mar 12 2003 clsnmpne
-rwxr-x---  1 root    system      55110 Mar 12 2003 pppsnmpd
lrwxrwxrwx  1 root    system      9 Apr 08 17:55 snmpd -> snmpdv3ne
```

而 hacmp 只支持 snmp version 1 . 所以我们要做一下调整：

```
stopsrc -s snmpd
/usr/sbin/snmpv3_ssw -1
startsrc -s snmpd
```

```

test2:/usr/sbin>ls -l |grep snmp
lrwxrwxrwx    1 root      system          18 Apr 21 13:40 clsnmp ->
/usr/sbin/clsnmpne
-rwxr-x---    1 root      system      83150 Mar 12 2003 clsnmpne
-rwxr-x---    1 root      system      55110 Mar 12 2003 pppsnmpd
lrwxrwxrwx    1 root      system          17 Apr 21 13:40 snmpd ->
/usr/sbin/snmpdv1

```

2. 启动 hacmp 时选择：

```

* Start now, on system restart or both [m          now
Start Cluster Services on these nodes          [test2]
BROADCAST message at startup?                true
Startup Cluster Lock Services?                false
Startup Cluster Information Daemon?            true
Reacquire resources after forced down ?        false

```

2. 执行 clstat

clstat - HACMP Cluster Status Monitor

Cluster: ha51tsc (1082085119)

Wed Apr 21 13:55:33 BEIDT 2004

State: UP

Nodes: 2

SubState: STABLE

Node: test1 State: UP

Interface: test1_boot1 (1) Address: 100.1.0.1

State: UP

Interface: test1_boot2 (1) Address: 192.168.0.1

State: UP

Interface: heartbeatdisk5 (0) Address: 0.0.0.0

State: UP

Interface: test1_svc (1) Address: 10.1.0.1

State: UP

Resource Group: res1 State: On line

Node: test2 State: UP

Interface: test2_boot1 (1) Address: 100.1.0.2

	State: UP
Interface: test2_boot2 (1)	Address: 192.168.0.2
	State: UP
Interface: heartbeatdisk7 (0)	Address: 0.0.0.0
	State: UP
Interface: test2_svc (1)	Address: 10.1.0.2
	State: UP
Resource Group: res2	State: On line