

HACMP : Modification d'un point de montage dépendant d'un Resource Group

Patrick DZIADKOWIAK

Le numéro 22 des Cahiers d'AIX, de janvier 1998, comprend un article traitant des nouveautés d'HACMP 4.2.1. Le chapitre 4 de cet article décrit les manipulations les plus courantes que l'administrateur HACMP doit réaliser. Ces instructions sont toujours valides avec les versions plus récentes d'HACMP.

Néanmoins, à l'époque il n'était pas possible de renommer un point de montage dépendant d'un Resource Group (tout au moins avec le cluster actif). Depuis la version 4.4.0 d'HACMP, ceci est désormais possible.

Avertissement

La modification d'un point de montage dépendant d'un *Resource Group* n'est pas possible en une seule opération de type DARE (*Dynamic Automatic Reconfiguration Events*). C'est-à-dire que l'on ne peut pas modifier le nom du point de montage dans le *Resource Group* et, directement, faire une synchronisation de *Resource Group* avec reconfiguration. Ceci se finirait inmanquablement par une magnifique erreur "*config_too_long*" sur l'événement "*reconfig_resource_release*"!

Ceci est dû au fait que l'odm d'HACMP utilise, comme référence, l'ancien point de montage qu'il essaye de démonter, alors que celui-ci a été renommé dans l'odm AIX ainsi que dans le fichier "*/etc/filesystems*".

Rappels sur ODM et HACMP

Pour mieux comprendre ce qui suit, il est nécessaire de se remémorer la “mécanique ODM” d’HACMP.

Copies de l’ODM

Il existe trois copies de l’ODM, chacune ayant une fonction spécialisée.

Première copie : DCD (*Default Configuration Data*)

Pour la personnalisation d’HACMP, les commandes appelées par **SMIT** travaillent dans les objets ODM du **DCD** (*Default Configuration Data*) du répertoire “/etc/objrepos”.

Deuxième copie : ACD (*Active Configuration Data*)

Depuis HACMP V4.2, quand on démarre HACMP sur un nœud, le DCD est recopié dans l’**ACD** (*Active Configuration Data*) du répertoire “/usr/sbin/cluster/etc/objrepos/active” ; tous les *daemons* actifs d’HACMP, ainsi que les utilitaires de “/usr/sbin/cluster/*” prennent leurs informations dans cet ODM.

Grâce à cette copie “active”, on peut faire des modifications sur la copie “par défaut”.

- Ce sont les opérations de synchronisation qui alignent ces deux ODM HACMP. Elles sont lancées, à la demande, par les sous-menus de “**smit**” :

SYNCHRONIZE CLUSTER TOPOLOGY
SYNCHRONIZE CLUSTER RESOURCES

Ces synchronisations sont obligatoires après chaque modification de topologie ou de ressource. Juste avant une opération de reconfiguration dynamique, l’ACD est sauvegardé, par “**snapshot**” dans un fichier “/usr/sbin/cluster/snapshots/active.X.odm”, où “X” est un chiffre de “0” à “9” avec “0” représentant toujours l’ACD le plus récent.

- En cas de problème durant l’opération de reconfiguration dynamique (*event reconfig FAILED*), il sera toujours possible de lancer le sous-menu de “**smit**” :

RELEASE LOCKS SET BY DYNAMIC AUTOMATIC RECONFIGURATION EVENT

Pour forcer le code retour à “0”, puis remplacer la sauvegarde active “.0.odm” dans le DCD avec le sous-menu de “**smit**” :

RESTORE DEFAULT CONFIGURATION FROM ACTIVE CONFIGURATION

Troisième copie : SCD (*Stage Configuration Data*) du répertoire

Pendant une opération de reconfiguration dynamique, le DCD de la machine initiatrice est recopié dans le DCD des autres nœuds du *cluster*, puis ce DCD est recopié dans un **ODM intermédiaire** : le **SCD** (*Stage Configuration Data*) du répertoire “/usr/sbin/cluster/etc/objrepos/stage”, avant que le SCD ne soit à son tour recopié dans l’ACD en même temps que les *daemons* HACMP sont rafraîchis pour prendre en compte le nouvel ACD.

Le SCD est alors détruit.

L'ODM d'AIX et l'ODM d'HACMP

Il ne faut surtout pas confondre :

- une modification de la configuration **HACMP** (topologie ou ressource), qui touche les objets ODM du DCD “/etc/objrepos/HACMP*”,
- avec les modifications d'un **LV**, qui concernent l'ODM d'AIX, ou les modifications d'un **FS** qui touchent “/etc/filesystems” et “CuAt” (si agrandissement de la taille du FS).

Pourtant, un FS (et son LV) sont, à la fois, une ressource HACMP et une unité AIX.

Synchronisation (rappels)

- La synchronisation des objets **ODM HACMP** se fait par :

```
# smitty hacmp
    Cluster configuration
        Cluster Resources
            Synchronize Cluster Resources

# smitty hacmp
    Cluster configuration
        Cluster Topology
            Synchronize Cluster Topology
```

- La synchronisation de l'ODM AIX se fait par :

```
....
exportvg
importvg
....
```

Méthode recommandée pour la modification d'un point de montage

Principe

Cette méthode consiste :

- à **retirer** le point de montage initial de la configuration HACMP,
- à faire une **synchronisation**,
- ensuite seulement, **renommer** le point de montage par CSPOC,
- **modifier** le *Resource Group*,
- et, enfin, à **synchroniser** de nouveau les *Cluster Resources*.

Prerequisite

Installer l'APAR **IY11638** :

PTF **U472532** pour 4.4.0.2 *ES Base Server Runtime*
PTF **U472534** pour 4.4.0.2 *HACMP Base Server Runtime*

Mise en œuvre

On trouvera, ci-dessous, le détail des panneaux à utiliser (toute l'opération s'effectue à partir d'un seul et même nœud ; CSPOC et DARE assurent l'homogénéité de la configuration entre les différents nœuds).

- Dans notre exemple, le point de montage “/proa_3” va être renommé en “proa_33”.

```
# smitty hacmp
  Cluster configuration
    Cluster Resources
      Change/Show Resources/Attributes for a Resource Group

[TOP]                                     [Entry Fields]
Resource Group Name                     proa_rg
Node Relationship                       cascading
Participating Node Names                proa prob

Service IP Label(s)                    [proa]
Filesystems                          <_1 /proa_2 /proa_3 /pr>
Filesystems Consistency Check          fsck
Filesystems Recovery Method            sequential
Filesystems/Directories to Export      [/proa_fsdir]
Filesystems/Directories to NFS Mount   [/proa_fsdirlocal;proa_>
Network For NFS Mount                  [] +
Volume Groups                          [proavg] +
Concurrent Volume Groups                [] +
Raw Disk PVIDs                         [] +
[MORE...11]
```

- On **retire** l'entrée “/proa_3” du champ “Filesystems” (DCD modifiée),
puis on **synchronise** les *Cluster Resources* (DCD + ACD mise à jour sur tous les nœuds + reconfiguration initiée)

```
# smitty hacmp
  Cluster configuration
    Cluster Resources
      Synchronize Cluster Resources

Ignore Cluster Verification Errors?    [No]
Un/Configure Cluster Resources?      [Yes]
* Emulate or Actual?                  [Actual]
* Skip Cluster Verification           [No]
```

- A l'issue, l'événement “*reconfig_resource_release*” est initié pour démonter “/proa_3”.
Vérifier le statut du *cluster* avant de poursuivre :
“/usr/sbin/cluster/clstat” doit annoncer un *cluster* stable.
- On **renomme** ensuite le point de montage au moyen de CSPOC, ce qui met à jour l'ODM AIX de l'ensemble des nœuds.

```
# smitty hacmp
Cluster System Management
Cluster Logical Volume Manager
Shared File Systems
Change / Show Characteristics of a Shared File System

[Entry Fields]
Resource Group Name      proa_rg
File system name         /proa_3
NEW mount point        [/proa_33]
SIZE of file system (in 512-byte blocks) [16384]
Mount GROUP              []
Mount AUTOMATICALLY at system restart?  no +
PERMISSIONS              read/write +
Mount OPTIONS            [] +
Start Disk Accounting?   no +
Fragment Size (bytes)    4096
Number of bytes per inode 4096
Compression algorithm     no
```

- On **édite** à nouveau le *Resource Group* pour ajouter le nouveau point de montage (DCD).

```
# smitty hacmp
Cluster configuration
Cluster Resources
Change/Show Resources/Attributes for a Resource Group

[TOP] [Entry Fields]
Resource Group Name      proa_rg
Node Relationship         cascading
Participating Node Names proa prob

Service IP Label(s)      [proa]
Filesystems           <_1 /proa_2 /proa_33 /pr>
Filesystems Consistency Check fsck
Filesystems Recovery Method sequential
Filesystems/Directories to Export [/proa_fsdir]
Filesystems/Directories to NFS Mount [/proa_fsdirlocal;proa_>
Network For NFS Mount    [] +
Volume Groups            [proavg] +
Concurrent Volume Groups [] +
Raw Disk PVIDs           [] +
[MORE...11]
```

- Enfin, on **synchronise** les *Cluster Resources* (DCD + ACD mise à jour sur tous les nœuds + reconfiguration initiée).

```
# smitty hacmp
  Cluster configuration
    Cluster Resources
      Synchronize Cluster Resources

Ignore Cluster Verification Errors? [Entry Fields]
Un/Configure Cluster Resources? [No]
* Emulate or Actual? [Yes]
* Skip Cluster Verification [Actual]
                                [No]
```

- A l'issue, l'événement "*reconfig_resource_release*" est initié pour monter "/proa_33".
Vérifier :
le statut du *cluster* : "/usr/sbin/cluster/clstat" (doit annoncer un *cluster* stable),
ainsi que "/tmp/hacmp.out" (vérifier qu'il n'y a pas d'erreur).

Variante de la méthode

Cette variante permet d'écourter l'indisponibilité du point de montage.

Il est possible, moyennant le "*démontage/montage manuels*", respectivement de l'ancien et du nouveau point de montage, de réduire à quelques secondes l'indisponibilité du point de montage pour les utilisateurs.

Cette méthode a été testée avec succès et, à l'issue du changement de nom du point de montage, des tests concluants de bascule ont été faits.

Prerequisite

Installer l'APAR **IY11638** :

PTF	U472532	pour 4.4.0.2 <i>ES Base Server Runtime</i>
PTF	U472534	pour 4.4.0.2 <i>HACMP Base Server Runtime</i>

Mise en œuvre

On trouvera ci-après le détail de cette procédure.

De la même manière que précédemment, toute l'opération s'effectue à partir d'un seul et même nœud.

- Ici encore, comme exemple, le point de montage "/proa_3" va être renommé en "/proa_33".
- On **renomme** le point de montage au moyen de CSPOC, ce qui met à jour l'ODM AIX de l'ensemble des nœuds.

```
# smitty hacmp
  Cluster System Management
    Cluster Logical Volume Manager
      Shared File Systems
        Change / Show Characteristics of a Shared File System

[Entry Fields]
Resource Group Name      proa_rg
File system name         /proa_3
NEW mount point        [/proa_33]
SIZE of file system (in 512-byte blocks) [16384]
Mount GROUP              []
Mount AUTOMATICALLY at system restart?  no      +
PERMISSIONS              read/write      +
Mount OPTIONS            []               +
Start Disk Accounting?   no               +
Fragment Size (bytes)    4096
Number of bytes per inode 4096
Compression algorithm     no
```

- On **édite** le *Resource Group* pour **renommer** le nouveau point de montage.

Remarque : N'avait-on pas dit, auparavant, qu'on ne pouvait pas le faire... ? Si ! Heureusement, il y a une subtilité...

```
# smitty hacmp
  Cluster configuration
    Cluster Resources
      Change/Show Resources/Attributes for a Resource Group

[TOP] [Entry Fields]
Resource Group Name      proa_rg
Node Relationship         cascading
Participating Node Names proa prob

Service IP Label(s)      [proa]
Filesystems             <_1 /proa_2 /proa_33 /pr>
Filesystems Consistency Check fsck
Filesystems Recovery Method sequential
Filesystems/Directories to Export [/proa_fsdir]
Filesystems/Directories to NFS Mount [/proa_fsdirlocal;proa_>
Network For NFS Mount    []      +
Volume Groups             [proavg]  +
Concurrent Volume Groups []      +
Raw Disk PVIDs           []      +
[MORE...11]
```

A ce niveau, seule la DCD est mise à jour.

- Maintenant, on va travailler à la place de HACMP pour **démonter/monter** les **ancien/nouveau** points de montage.

```
# umount /proa_3
# mount /proa_33
```

Le montage de “/proa_33” est possible car l’ODM AIX ainsi que “/etc/filesystems” sont à jour avec le nouveau nom du point de montage.

- Maintenant, la **véritable astuce** consiste à **mettre à jour** l’ODM active d’HACMP (ACD) par une **synchronisation** des *Cluster Resources*, **sans actualiser** “clstrmgr” (sans reconfiguration).

En effet, comme le *cluster* est déjà dans la nouvelle situation, il ne faut pas actualiser “clstrmgr” qui chercherait d’abord à démonter l’ancien point de montage. Ceci aurait pour conséquence, comme nous l’avons déjà expliqué, de générer une erreur “config_too_long”.

On actualise donc **uniquement** l’ODM HACMP (ACD) ce qui permettra, lors du prochain événement (exemples : “move resource”, “node_down”), de ne plus avoir en référence l’ancien point de montage mais seulement le nouveau.

```
# smitty hacmp
    Cluster configuration
        Cluster Resources
            Synchronize Cluster Resources

Ignore Cluster Verification Errors? [Entry Fields]
Un/Configure Cluster Resources? [No]
* Emulate or Actual? [Actual]
* Skip Cluster Verification [No]
```

A l’issue de la synchronisation, aucun événement n’est généré. Donc “clstat” annonce le *cluster* toujours en “état stable” et il n’y a pas d’activité dans “hacmp.out”.

