Enterprise Linux 實戰講座－RHEL High-Availability Solution（三）

High Availability SAMBA Cluster on SAN

由於 SAN 的當紅，而且愈來愈多廠商對於 Linux HA 解決方案感興趣，本期文章筆者將介紹如何在 FC SAN 環境下實作 RedHat High Availability SAMBA Cluster。

簡介

儲存區域網路（Storage Area Network）簡稱 SAN，是一種連結儲存裝置的專屬網路，其目的是用來取代伺服器和儲存設備之間的 SCSI I/O。提到 SAN，絕大多數人會將它和光纖通道（Fibre Channel）聯想在一起，事實上根據 SNIA（Storage Networking Industry Association，SNIA）的定義，SAN 是指專爲儲存設計的網路架構，並不限定使用何種網路技術，光纖通道技術只是儲存網路技術的其中一項。

雖然目前絕大多數的 SAN 都採用光纖通道技術建構儲存網路，但除了光纖通道技術外，SAN 也可以採用乙太網路技術建構。SNIA 建議，以光纖通道技術建構的儲存網路稱爲 Fibre Channel SAN（FC SAN），以乙太網路技術（如 iSCSI）建構的儲存網路則稱爲 IP SAN。

SAN 的資料傳輸走的是專用網路，而 FC SAN 所採用的光纖通道傳輸協定的頻寬高達 2Gbit/s），和 SCSI 匯流排的傳輸速率差不多，除了適合存放需要運算的資料（如資料庫），尤其適合應用在資料的備份與還原，而且不僅不會增加區域網路的流量負擔，亦不佔用伺服器的運算資源。

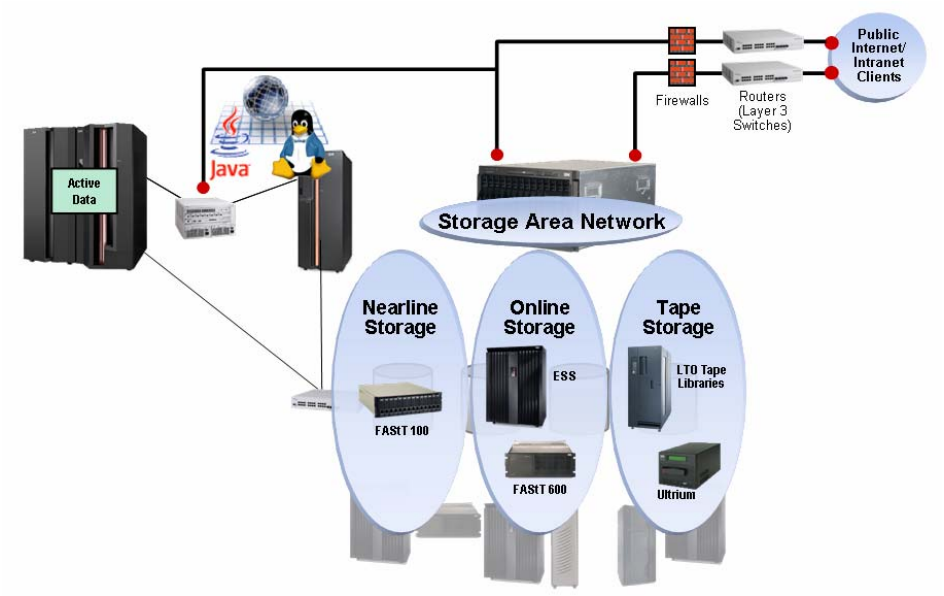

圖 1：企業 SAN 架構示意圖

前期文章，筆者利用 Proware Rackmount 3000I3 磁碟陣列當 HA Cluster 所需的 Share Disk。在 SCSI 的架構下，Linux Cluster 會有諸多的限制，例如 node 數的擴充，會受到 SCSI 磁碟陣列所能串接的主機數限制（通常爲兩台）及 SCSI 排線長度限制。反觀若採用 SAN 的架構就不會如此綁手綁腳。

SAMBA 是讓 Linux 與 MS Windows 共享資源的服務。讓 Windows 電腦可以透過『網路上的芳鄰』來存取 Linux 主機上面的檔案。通常扮演企業中 File Server 的角色。此篇文章最主要便是帶領各位讀者如何在 SAN 架構下建置 RedHat High Availability SAMBA Cluster。

測試環境

軟體

- **RedHat Enterprise Linux AS 版 Update 1**
- **RedHat Cluster Suite Update 2**

硬體

- x 86 伺服器兩台
- IBM FAStT 500/EXP 500
- SAN Switch
- IBM FAStT FC-2 Host Bus Adapter（QLA2300）兩張
- 兩張網路卡

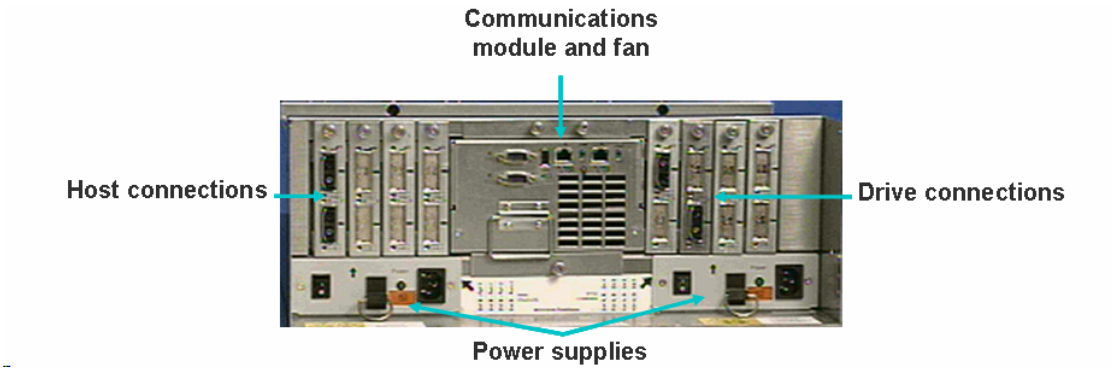

圖 2：IBM FAStT 500 背面架構圖

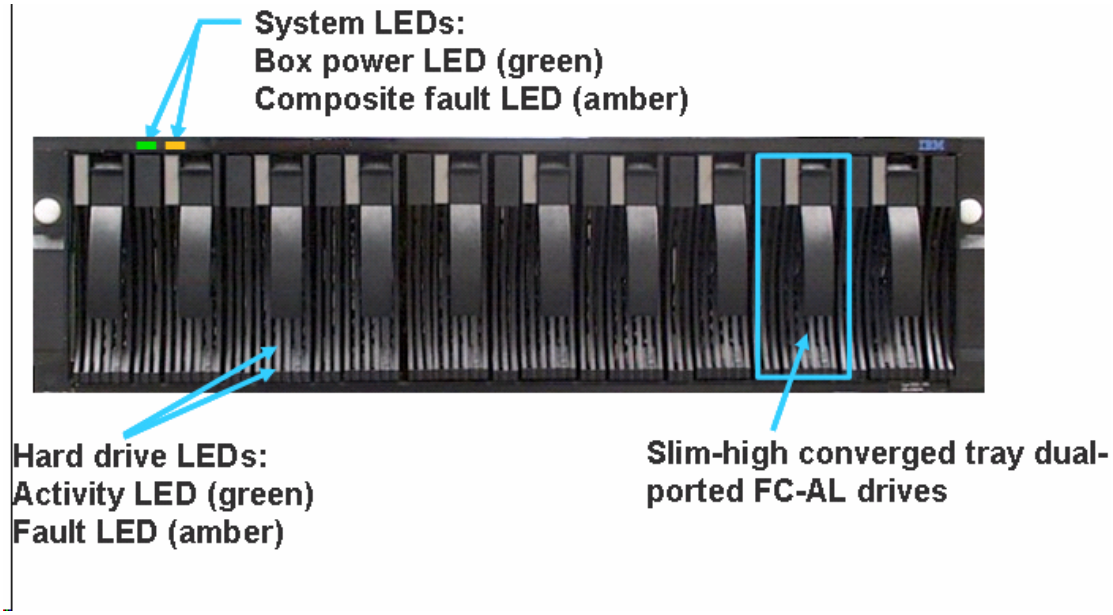

圖 3：IBM EXP500 Storage Expansion Unit 正面圖

圖 4：IBM 3534-F08 Fibre Channel Switch

圖 5：IBM FAStT FC-2 Host Bus Adapter（QLA2300）HBA 卡

High Availability SAMBA Cluster on SAN 架構

筆者測試架構的簡圖如圖 6。主要伺服器 rhel3-1 的 ip 爲 192.168.33.1，備援伺服器 rhel3-2 的 ip 爲 192.168.33. 2，整個 HA Cluster 對外的 service ip 爲 192.168.33.3。

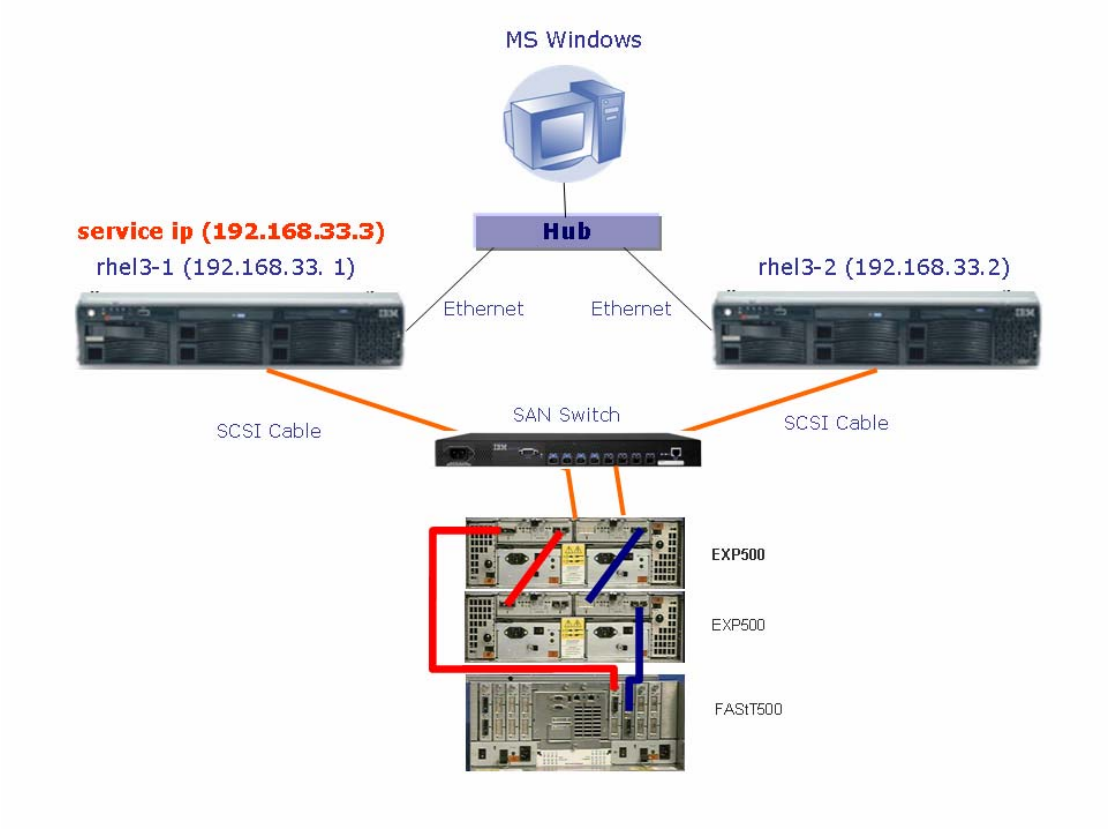

圖 6：High Availability SAMBA Cluster on SAN 架構圖

Linux SAMBA HA Cluster on SAN 的建置分爲三個部份

一、Setup fibre channel storage

1.設定Qlogic HBA卡
2.Update Qlogic driver
3.Create New Logical Drive
4.設定Mappings

二、Install and setup Red Hat Cluster Manager

1.安裝Red Hat叢集管理員套件
2.編寫/etc/hosts
3.設定叢集共用分割區
4.訂定叢集名稱
5.設定Share Raw Device
6.新增Cluster Member
7.設定Failover Domain
8.SAMBA Druid
9.複製相關設定檔至另一台node

10.設定Cluster Log
11.查看Cluster狀態

三、測試

一、Setup fibre channel storage

1.設定Qlogic HBA卡

當系統開機時，出現 Qlogic BIOS 提示符號時，按下「**Ctrl+Q**」
除了修改下列設定值外，其餘皆保留預設值（圖 7）：

- Host Adapter settings
Loop reset delay - 8.
- Advanced Adapter Settings
LUNs per target - 0
Enable Target Reset - Yes
Port down retry count - 12

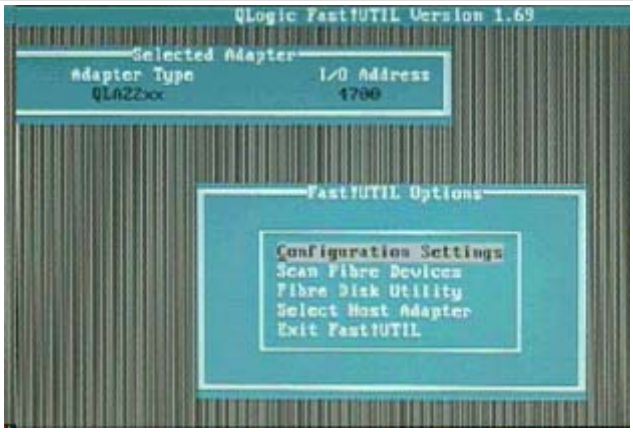

圖 7：Fast!UTIL 設定畫面

2.Update Qlogic driver

雖說筆者測試過 RedHat Enterprise U2 內附的 Qlogic Driver 是可以正確偵測到 FAStT 分配的磁碟空間。不過筆者還是建議使用 IBM 所提供的 Qlogic Driver，畢竟這是 IBM Lab 測試過且認爲沒有問題的版本。筆者使用的是 IBM FAStT Host Adapter Driver 7.00.61 版，此版本的 driver 可從下列網址 download。

http://www-1.ibm.com/support/docview.wss?rs=501&uid=psg1MIGR-54952&loc=en_US

它所支援的 Linux O.S 如表一所示。就筆者的經驗，如果在 Linux 上安裝 HBA 卡，一定要注意 kernel 版本和 Driver 版本搭配相容問題，根據表一，7.00.61 版支援 RHEL3 的 kernel 版本爲 2.4.21.9.0.1.EL，所以筆者的實作環境便採用 2.4.21.9.0.1.EL 的 kernel。

表 1：IBM FAStT Host Adapter Driver 7.00.61 版支援作業系統列表

作業系統	Kernel 版本
RedHat Advanced Server 2.1 32-bit	2.4.9-e.40 UP, SMP, Enterprise(Bigmem) and Summit
RedHat Advanced Server 2.1 IA64 (Homogeneous Only)	2.4.9-e.41 UP, SMP, Enterprise(Bigmem) and Summit
RedHat 3.0 IA-32	2.4.21-9.0.1 EL
RedHat 3.0 IA=64 (Homogeneous Only)	2.4.21.9.0.1.EL
United Linux 1.0 with SP3 32-bit	2.4.21-198
United Linux 1.0 IA-64 (Homogeneous Only)	2.4.21-203
United Linux 1.0 AMD-64 (Homogeneous Only)	2.4.21-212

■ 安裝 2.4.21-9.0.1 EL kernel 及 source

若是讀者安裝的是 RHEL 3.0，預設的 kernel 爲 2.4.21-4，則必須跟 RedHat 申請 Update 1 CD，在此光碟中 RedHat/Updates 目錄中包含 2.4.21-9.0.1.EL 的 kernel 及 kernel source。請利用 rpm 指令安裝，並重新開機，選擇從 2.4.21-9.0.1.EL kernel 開機。如果讀者安裝的是 RHEL 3.0 Update 1 則其預設的 kernel 便是 2.4.21-9.0.1.EL，所以只要再安裝 kernel source 的 rpm 即可。

```
#rpm -ivh kernel-2.4.21-9.EL.i686.rpm
#rpm -ivh kernel-source-2.4.21-9.EL.i386.rpm
```

■ 解壓縮 Qlogic Driver Source code

```
#tar -xvzf 26r0536.tgz –C /tmp
#cd /tmp/i2x00-v7.00.61
#./drvrinstall
```

■ Build Qlogic device driver

#cd /usr/src/linux-2.4

#vi Makefile 將 EXTRAVERSION 中的 custom 刪除

```
VERSION = 2
PATCHLEVEL = 4
SUBLEVEL = 21
EXTRAVERSION = -9.0.1.EL~~custom~~
```

#make mrproper
#cp configs/kernel-2.4.21-i686.config .config
註：若讀者的機器爲多處理器（smp）架構，則需要使用 kernel-2.4.21-i686-smp.config
#make oldconfig
#make dep

切換到 Qlogic Source Code 的目錄
#cd /tmp/i2x00-v7.00.61

利用 make all 產生 Qlogic Driver，若是 smp 架構，則用 make all smp=1，執行此道指令後，在此目錄下會產生 Qlogic 2200 及 Qlogic 2300 HBA 卡的 Driver。

#make all
#cp *.o /lib/modules/2.4.21-9.EL/kernel/drivers/addon/qla2200/
#vi /etc/modules.conf
在檔案最後加上一行
options scsi_mod max_scsi_luns=32
#reboot
重新開機

3.Create New Logical Drive

筆者欲從 FAStT 500 上切出一塊 1 GB 的空間分配給 Linux HA Cluster，首先必須安裝 FAStT Storage 管理工具 IBM FAStT Storage Manager v8.4，下載網址如下：

```
http://www-1.ibm.com/support/docview.wss?rs=501&uid=psg1MIGR-52951&loc=en_US
```

筆者可以將其安裝在 Windows 或 Linux 上，啓動 Storage Manager 所看到的畫面如圖 8。

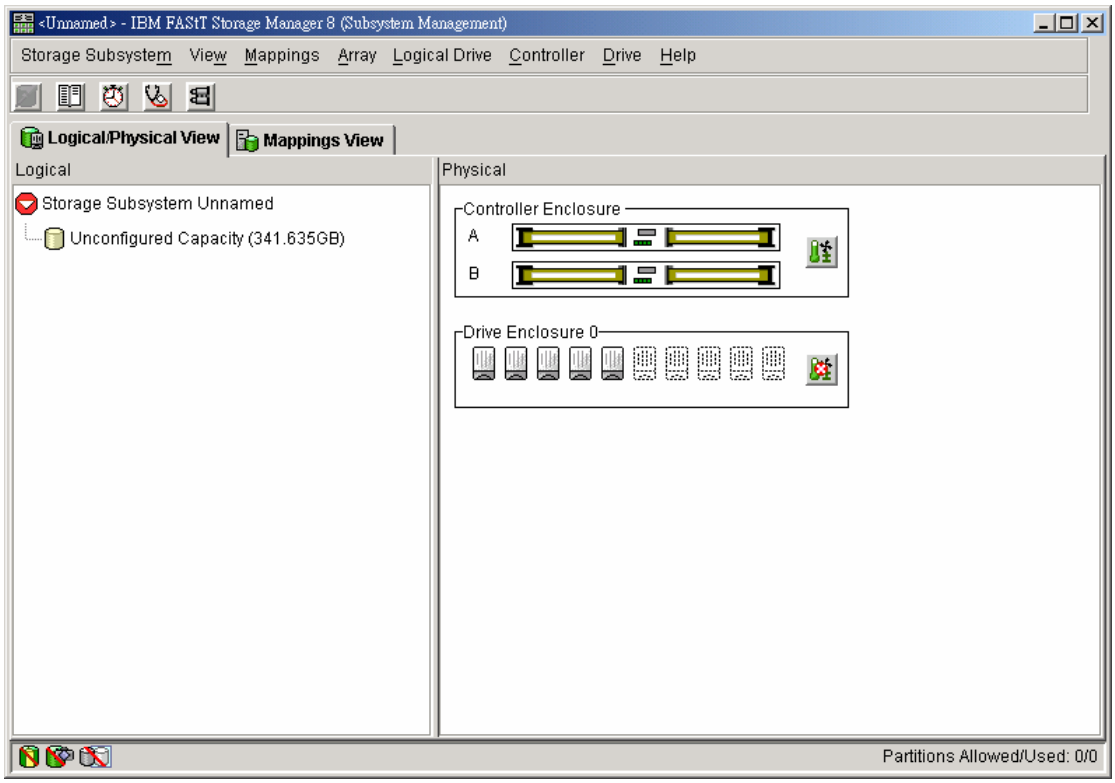

圖 8：IBM FAStT Storage Manager v8.4 畫面

■ 點選「**Unconfigured Capacity**」，然後按「**右鍵**」選取「**Create New Logical Drive**」

■「**Current default host type**」展開下拉式選單，選取「**linux**」

■ 此時會出現「**Create Logical Drive Wizard**」畫面，請選取「**Unconfigured Capacity (create new array)**」然後按下「**Next**」。（圖 9）

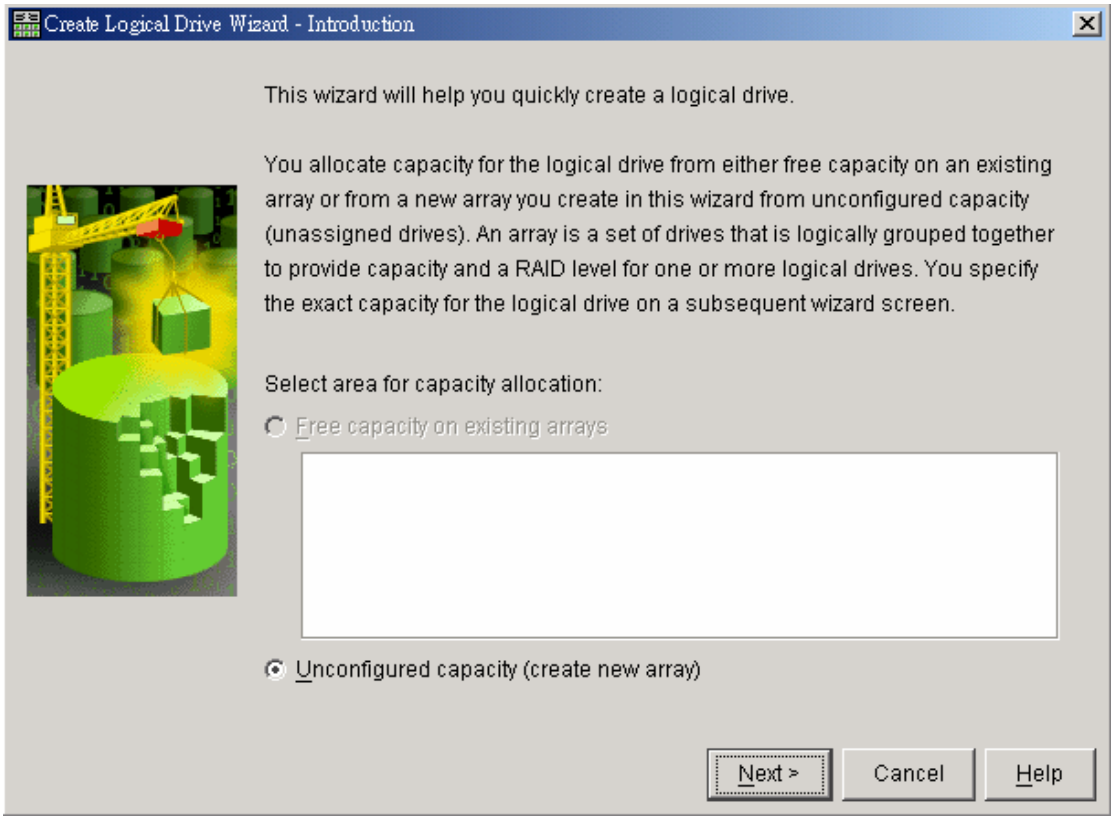

圖 9：Create Logical Drive Wizard

■ 選取適當的 RAID level，筆者設定爲「**RAID 5**」

■ 讀者可以利用自動（Automatic）或手動（Manual）的方式選取欲利用那些硬碟組成 RAID 5 的 Disk Array。（圖 10）

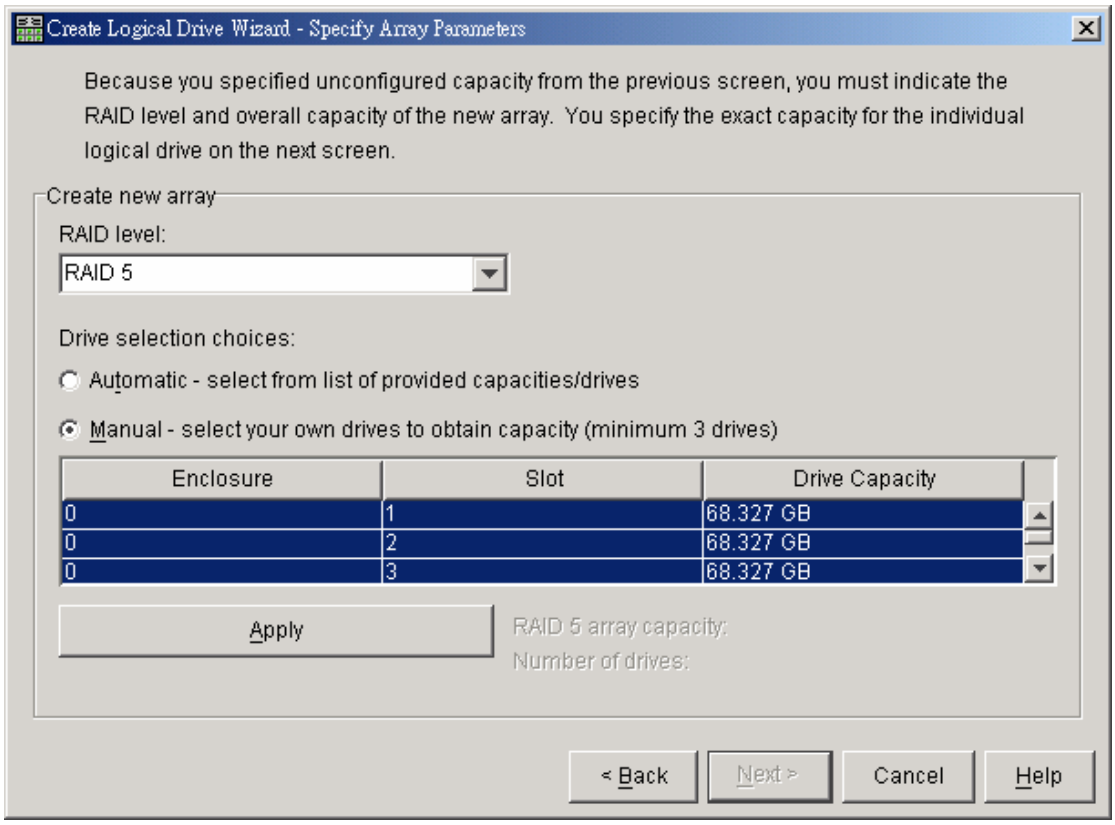

圖 10：指定 Disk Array 相關參數

■ 設定 New logical drive 的大小（圖 11）

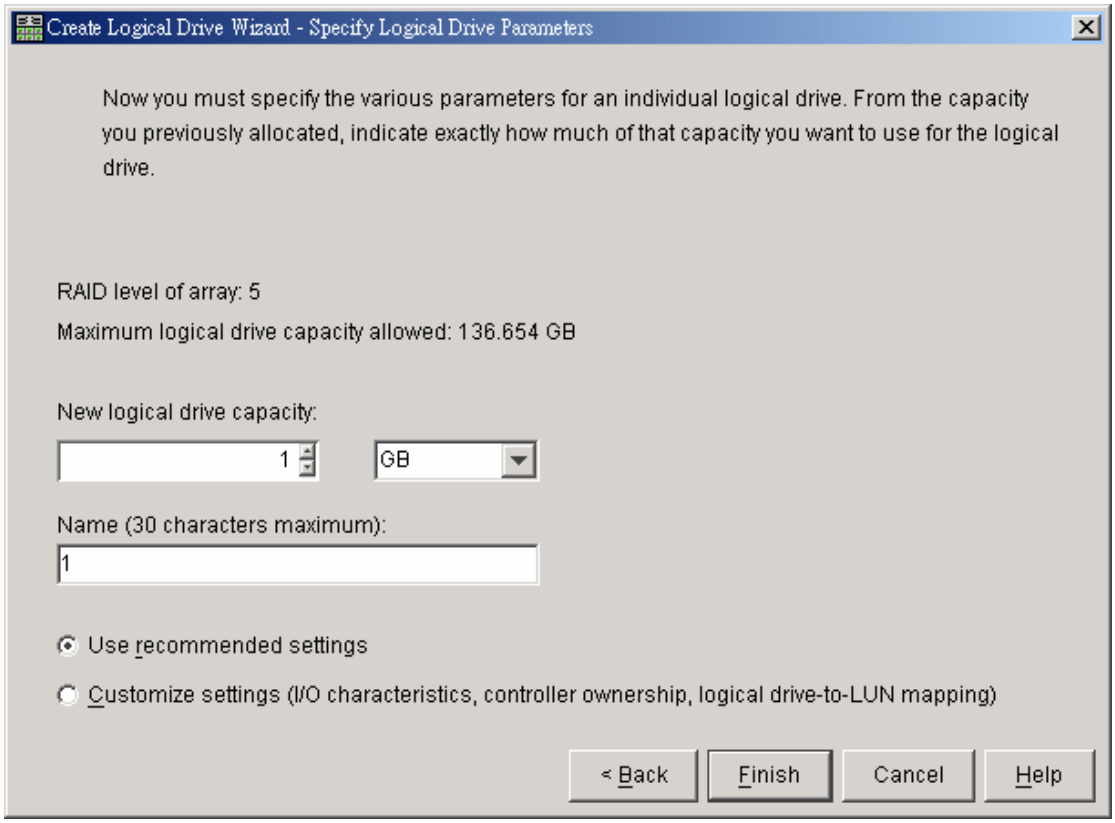

圖 11：設定 New logical drive 的大小

4.設定Mappings

■ 選取FAStT Storage Manager 8主畫面「Mappings View」，點選「Default Group」按右鍵，選取「Define Host」新增兩個Host。（圖12）

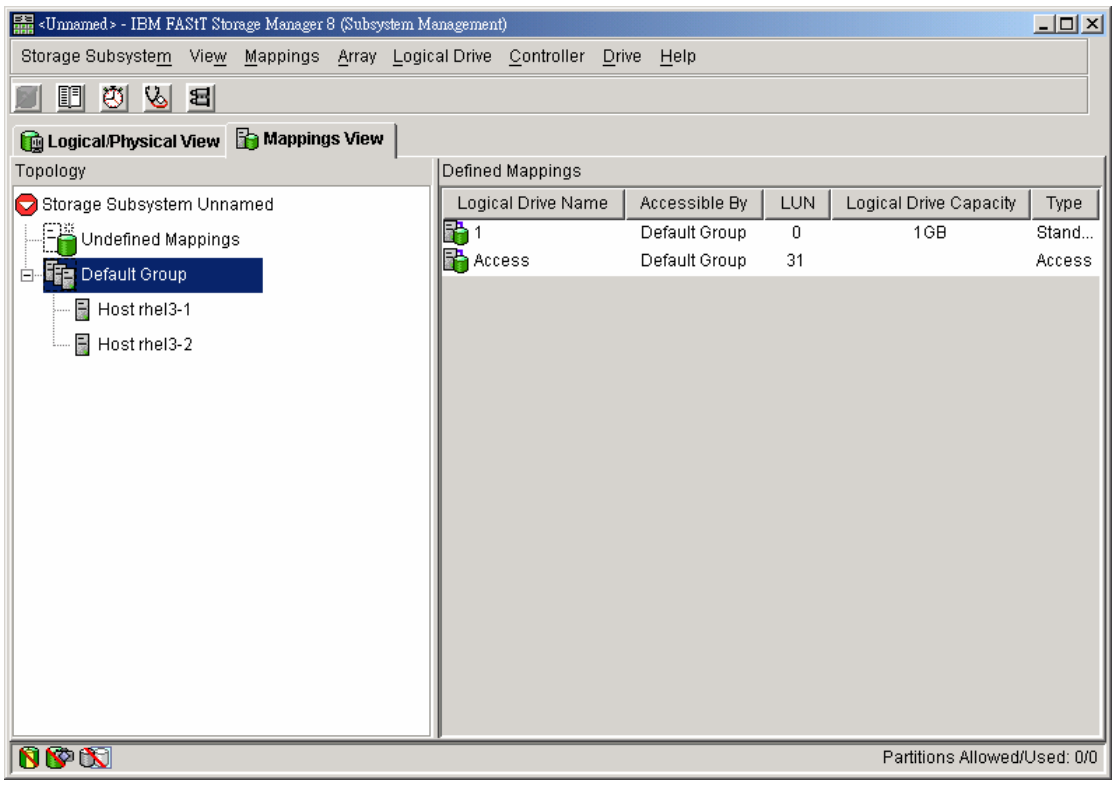

圖 12：「Mappings View」定義 Host 畫面

■ 點選右邊視窗中的「**Logical Drive Name**」，按右鍵選取「**Change Mapping**」。（圖 13）

■ 在 Define Host Port 畫面中填入 rhel3-1 主機上 HBA 卡的 Host port identifier（圖 14）；重覆此步驟定義另一台主機 rhel3-2。最後會看到如圖 15 的畫面。

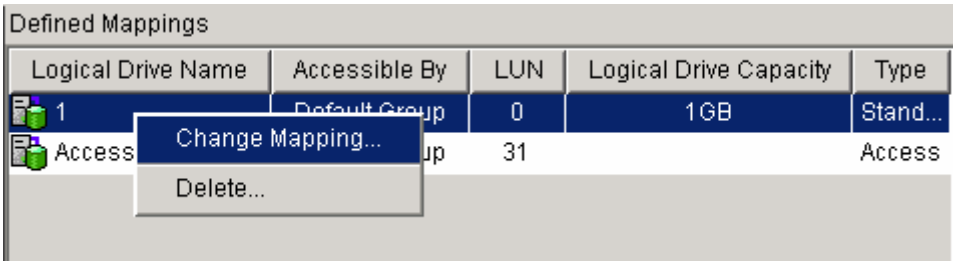

圖 13：Defined Mappings 畫面

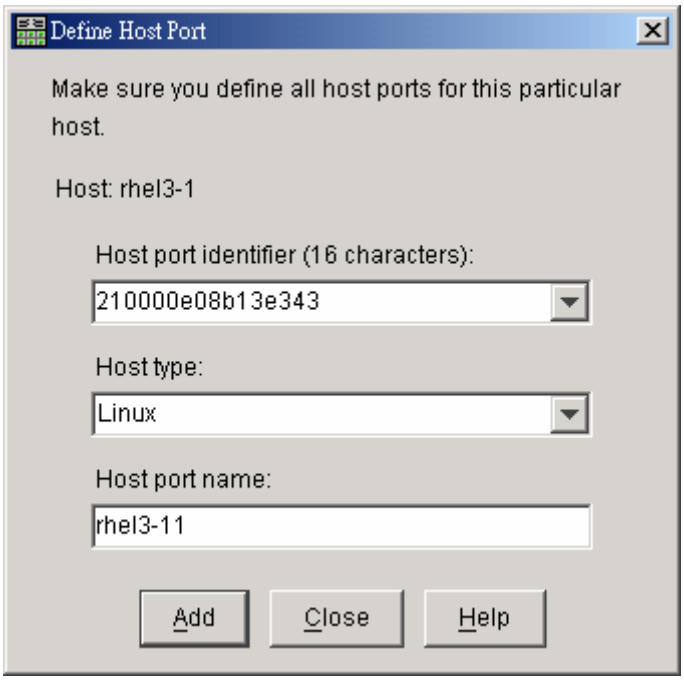

圖 14：Defined Host Port 畫面

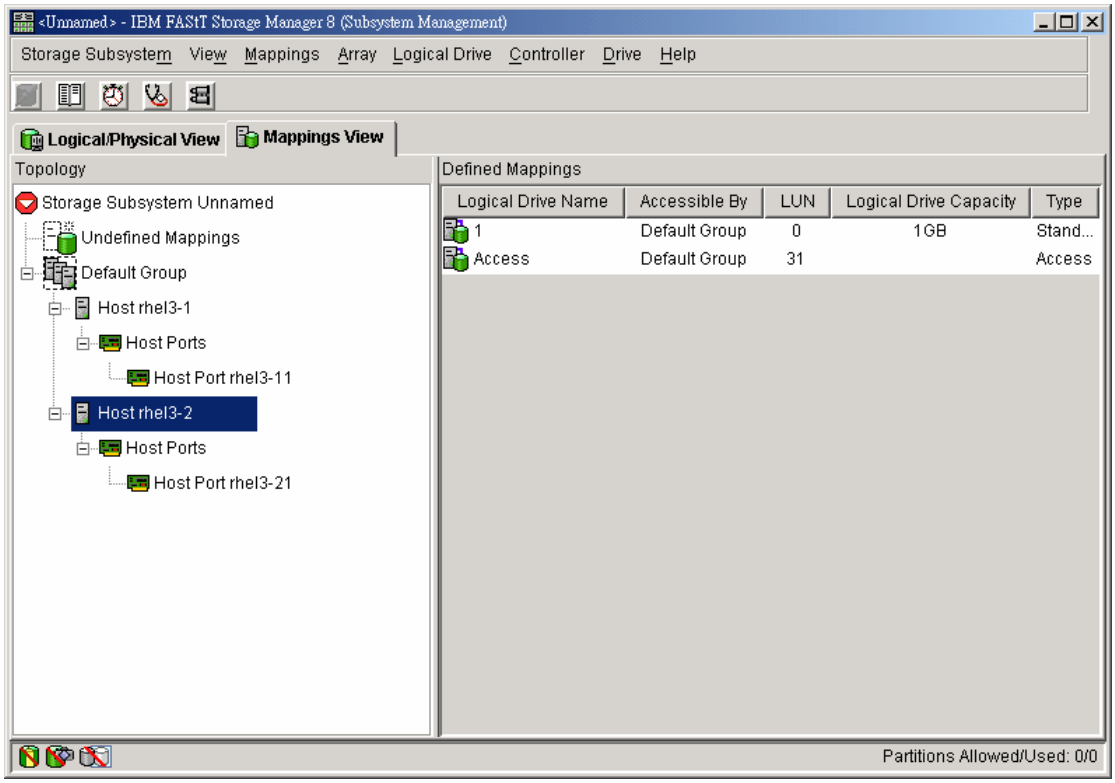

圖 15：Mappings View 設定完成畫面

二、Install and setup Red Hat Cluster Manager

1.安裝 Red Hat 叢集管理員套件

用 root 登入 rhel3-1 安裝 clumanager 與 redhat-config-cluster 套件才能設定 Red

Hat 叢集管理員，將光碟收入光碟機中，便會自動執行安裝程式。請選取「clumanager」及「redhat-config-cluster」套件進行安裝；在 rhel3-2 亦重複此步驟，或利用 rpm 方式安裝：

```
# rpm -ivh clumanager-1.2.12-1.i386.rpm
# rpm -ivh redhat-config-cluster-1.0.2-1.1.noarch.rpm
```

2.編寫 /etc/hosts

```
[root@rhel3-1 root]# cat /etc/hosts
# Do not remove the following line, or various programs
# that require network functionality will fail.
127.0.0.1       localhost.localdomain localhost
192.168.33.1 rhel3-1.example.com rhel3-1
192.168.33.2 rhel3-2.example.com rhel3-2
192.168.33.3 ha.example.com          ha
[root@rhel3-1 root]# scp /etc/hosts rhel3-2:/etc 並將此檔 scp 至 rhel3-2
```

3.設定叢集共用分割區（Configuring Shared Cluster Partitions）

■ 在 rhel3-1 上利用 fdisk 切出兩個 60MB 的分割區 sdb1、sdb2（須大於 10MB）做 Raw Device 用，再從其中切出 500MB 分割區，再利用「**mke2fs –j /dev/sdb3**」格式化此檔案系統。（圖 16）

■ 編寫/etc/sysconfig/rawdevices 檔案，編輯完/etc/sysconfig/rawdevices 檔案後，可以重新開機 或者是執行下列指令「**service rawdevices restart**」來使其生效。

■ 讀者可以用「**raw –aq**」查詢所有的 raw 裝置

```
[root@rhel3-1 root]# cat /etc/sysconfig/rawdevices
# raw device bindings
# format:   <rawdev> <major> <minor>
#                <rawdev> <blockdev>
# example: /dev/raw/raw1 /dev/sda1
#                /dev/raw/raw2 8 5
/dev/raw/raw1 /dev/sdb1
/dev/raw/raw2 /dev/sdb2
```

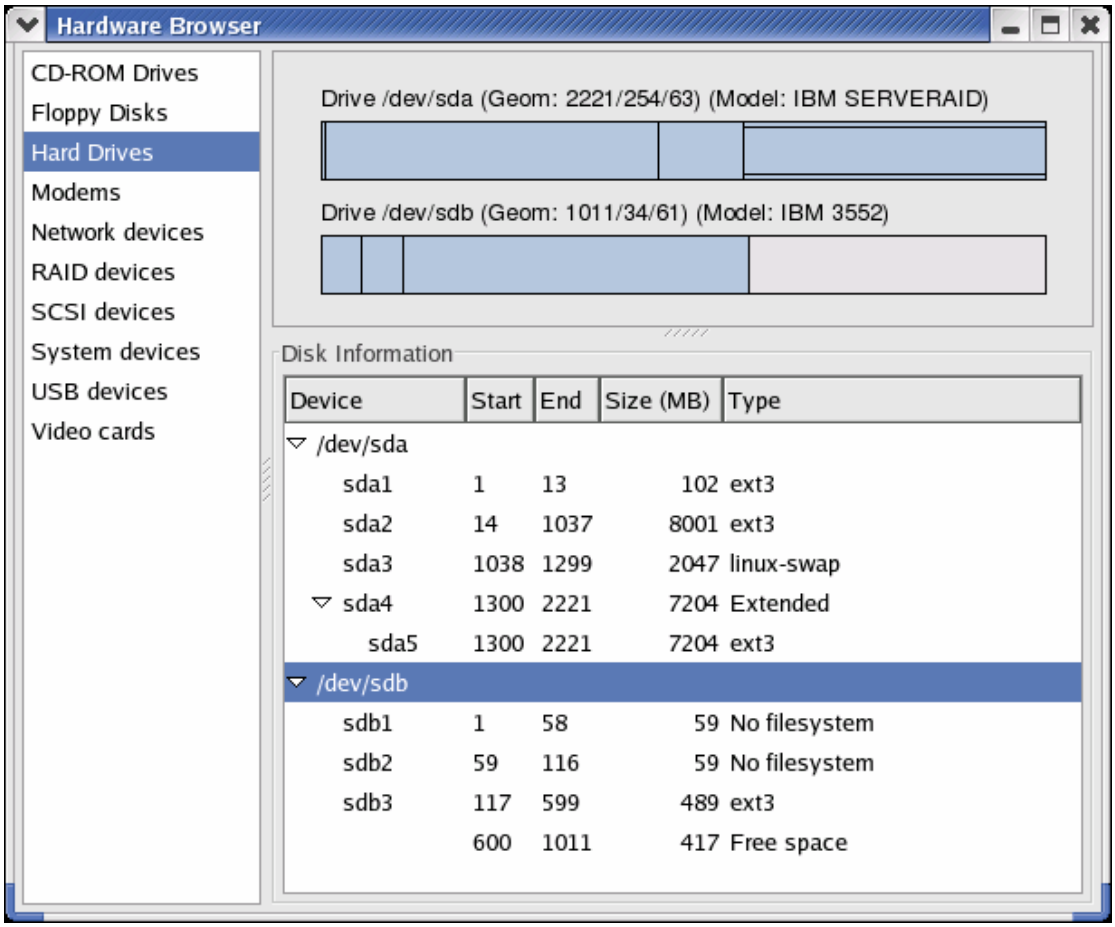

圖 16：hwbrowser 顯示 share disk 分割狀況

4.訂定叢集名稱

- 選擇『**主選單**』/『**系統設定**』/「**伺服器設定**』/『**叢集**』。
- 或在 shell 提示符號下輸入 redhat-config-cluster 指令。

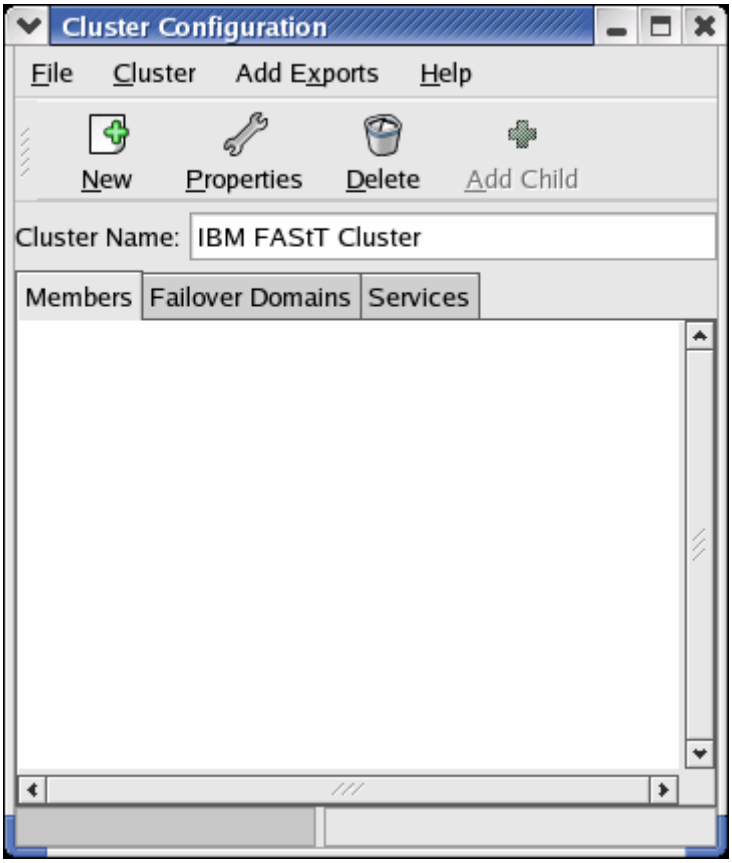

圖 17：訂定叢集名稱

5.設定 Share Raw Device

選擇 redhat-config-cluster 上的「**Cluster**」/「**Shared State**」便可看到圖 18 的畫面，填入正確的 Raw Device。

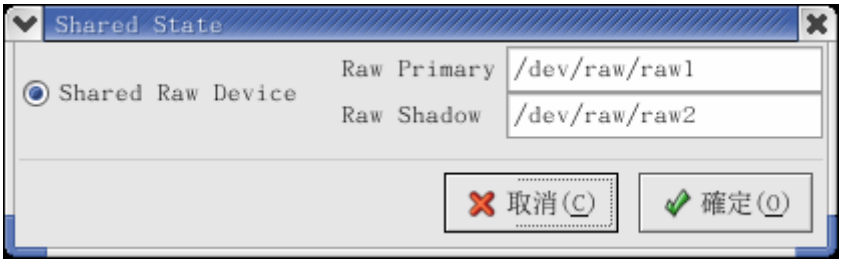

圖 18：Share Raw Device

6.新增 Cluster Member

選取「**Member**」，再點選「**New**」的按鈕。便會出現要求輸入 Member 名稱的視窗。請輸入 Cluster 中一部系統的主機名稱或位址，請注意每一個 Member 必須位於與執行 redhat-config-cluster 的機器在同一子網路中，而且必須在 DNS 或每一部叢集系統的 /etc/hosts 檔案中已經定義了。請新增兩個 Cluster Member「rhel3-1」及「rhel3-2」。

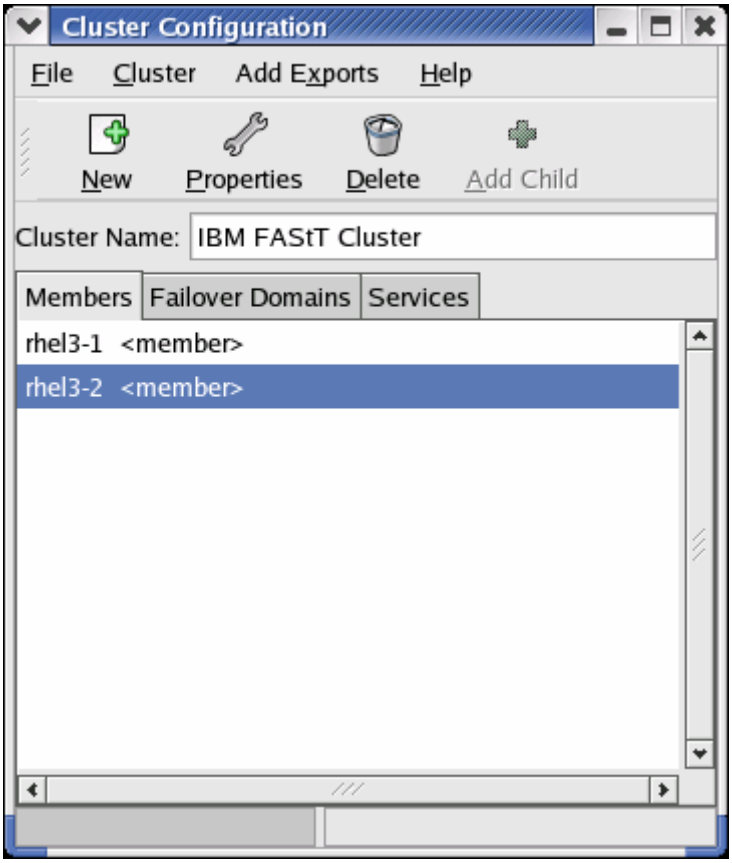

圖 19：Linux HA cluster Members

7.設定 Failover Domain

選擇「**Failover Domain**」的標籤頁，再點選「**New**」的按鈕。將會出現的「**Failover Domain**」對話視窗，設定 Domain Name 及 Add Members。

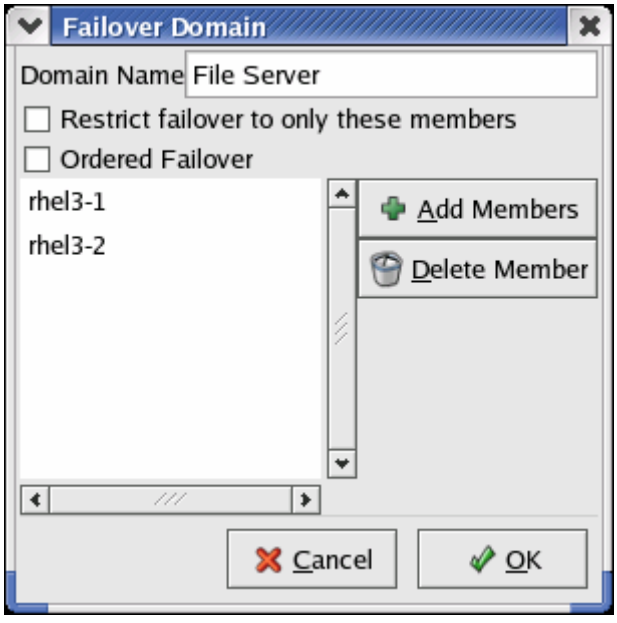

圖 20：設定「Failover Domain」的屬性

8.利用「SAMBA Druid」來快速設定一個用戶端可存取的 SAMBA 共享

■ 點選「**Add Exports**」/「**SAMBA**」，將會看到如圖 21 的畫面，然後按下「**Forward**」。

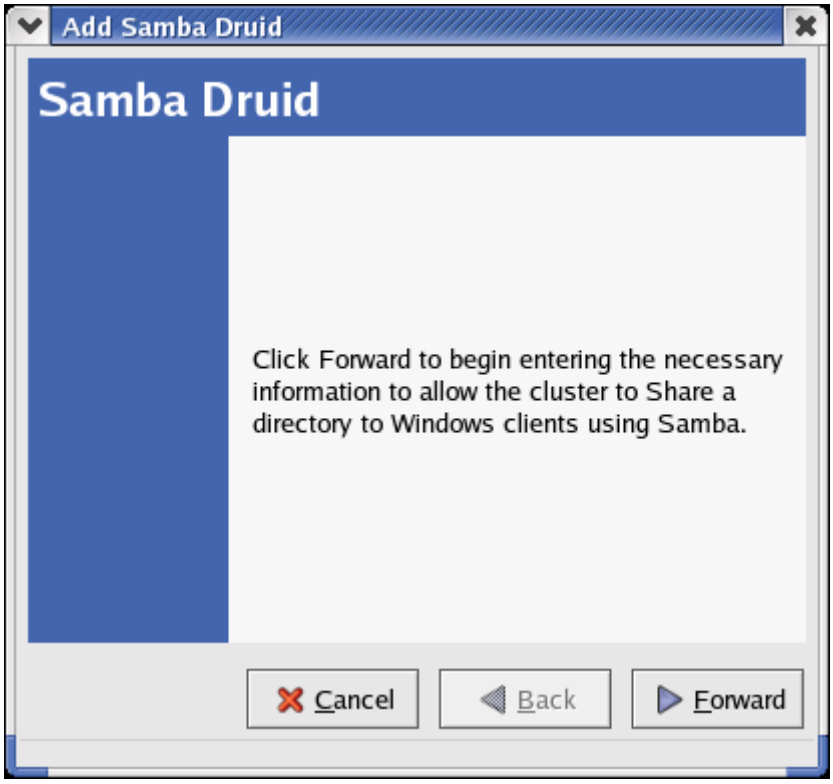

圖 21：SAMBA Druid 畫面

■ 利用「**SAMBA Druid**」將/dev/sdb3 分享給 Windows Client，首先設定「Service Name」及「Service IP」。（圖 22）

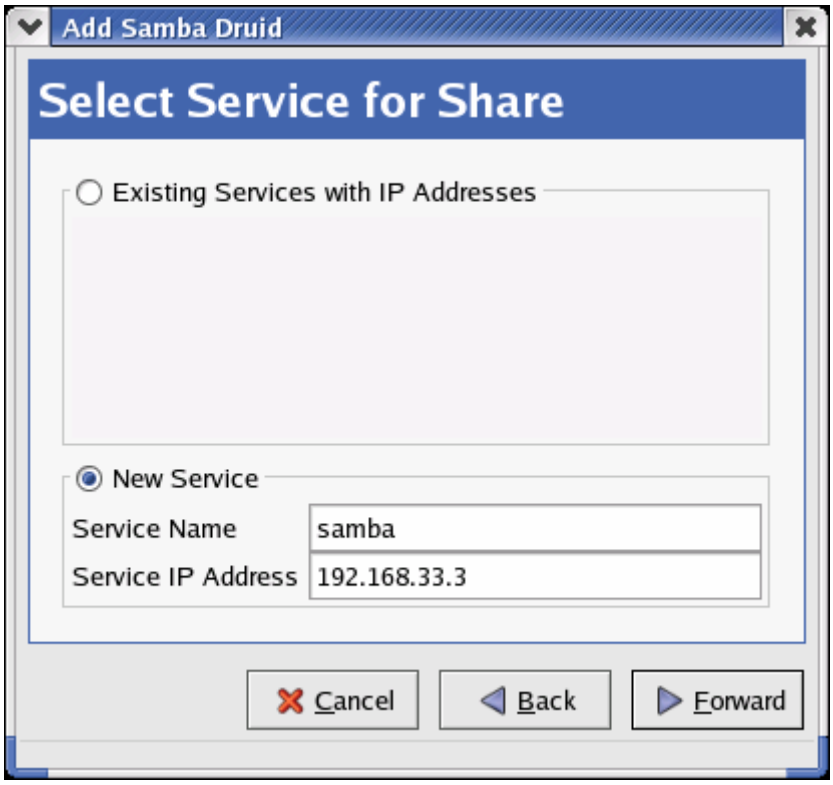

圖 22：設定 Service Name 及 Service IP Address

■ 設定 SAMBA service 所對應的 Device Special File 及 Device Mount Point 如圖 23 所示。按下「**Forward**」鍵後會出現設定 share name 的視窗（圖 24）。

圖 23：設定欲分享的 Device 及對應的 Mountpoint

圖 24：設定 Share Name

■ 設定 Share Name 後，按下「**Forward**」鍵，出現設成畫面。記得將 /etc/samba/smb.conf.myshare 複製至其他的 Cluster Memeber。

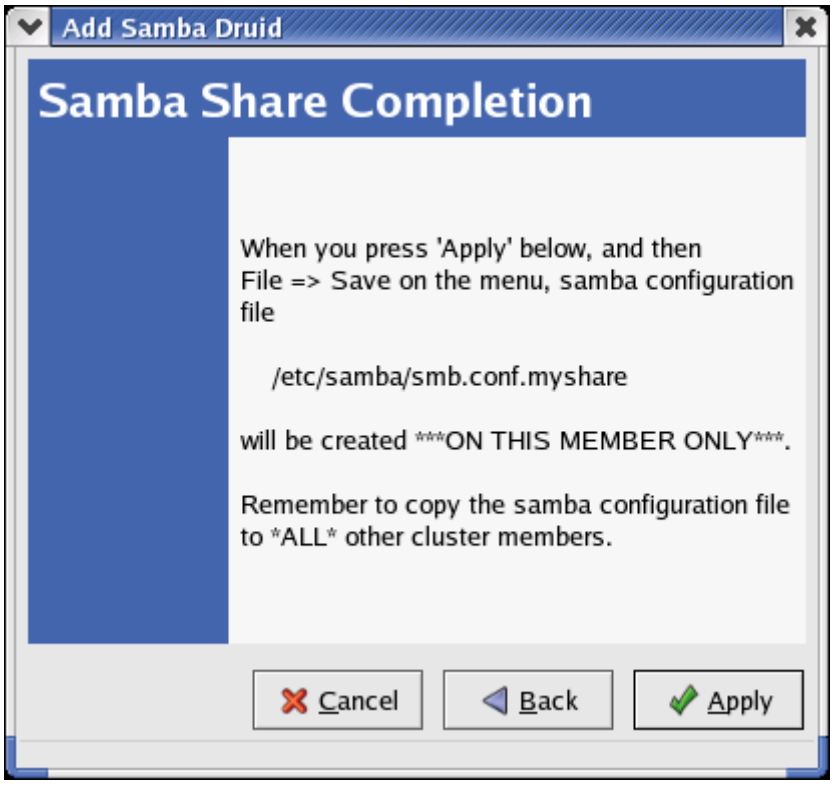

圖 25：SAMBA Druid 設定完成畫面

■ 最後點選「**Apply**」完成這個服務。並從「**叢集設定工具**」中選擇「**File**」/「**Save**」來儲存設定。(圖 25)

9.複製相關設定檔至另一台 node

將相關設定檔 cluster.xml、/etc/samba/smb.conf.myshare 複製至 rhel3-2

```
[root@rehl3-1 root]# scp /etc/cluster.xml rhel3-2:/etc/
root@rhel3-2's password:
cluster.xml                                   100% 1405
```

```
[root@rehl3-1 root]# scp /etc/samba/smb.conf.myshare rhel3-2:/etc/
root@rhel3-2's password:
smb.conf.myshare                              100% 867
```

■ 在 rhel3-1 及 rhel3-2 上啓動 clumanager Daemon

```
[root@ rehl3-1 root]# service clumanager start
Starting Red Hat Cluster Manager...
Loading Watchdog Timer (softdog):                    [  OK  ]
Starting Quorum Daemon:
```

```
[root@ rehl3-2 root]# service clumanager start
Starting Red Hat Cluster Manager...
Loading Watchdog Timer (softdog):                          [   OK   ]
Starting Quorum Daemon:
```

10.設定 Cluster log

修改 rhel3-1 及 rhel302 上的/etc/syslog.conf 指定 Cluster Log 存放位置，並重新啓動 syslog Daemon。

```
[root@rhel3-1 root]# vi /etc/syslog.conf
# Add for cluster
local4.*              /var/log/cluster

# service syslog restart
```

```
[root@rhel3-2 root]# vi /etc/syslog.conf
# Add for cluster
local4.*              /var/log/cluster

# service syslog restart
```

11.查看 Cluster 狀態

在終端視窗中鍵入「**redhat-config-cluster**」啓動 Cluster 管理工具。點選視窗下方 samba service，然後勾選啓動 samba 服務。如果一切設定無誤，應會看到如圖 25 的畫面。

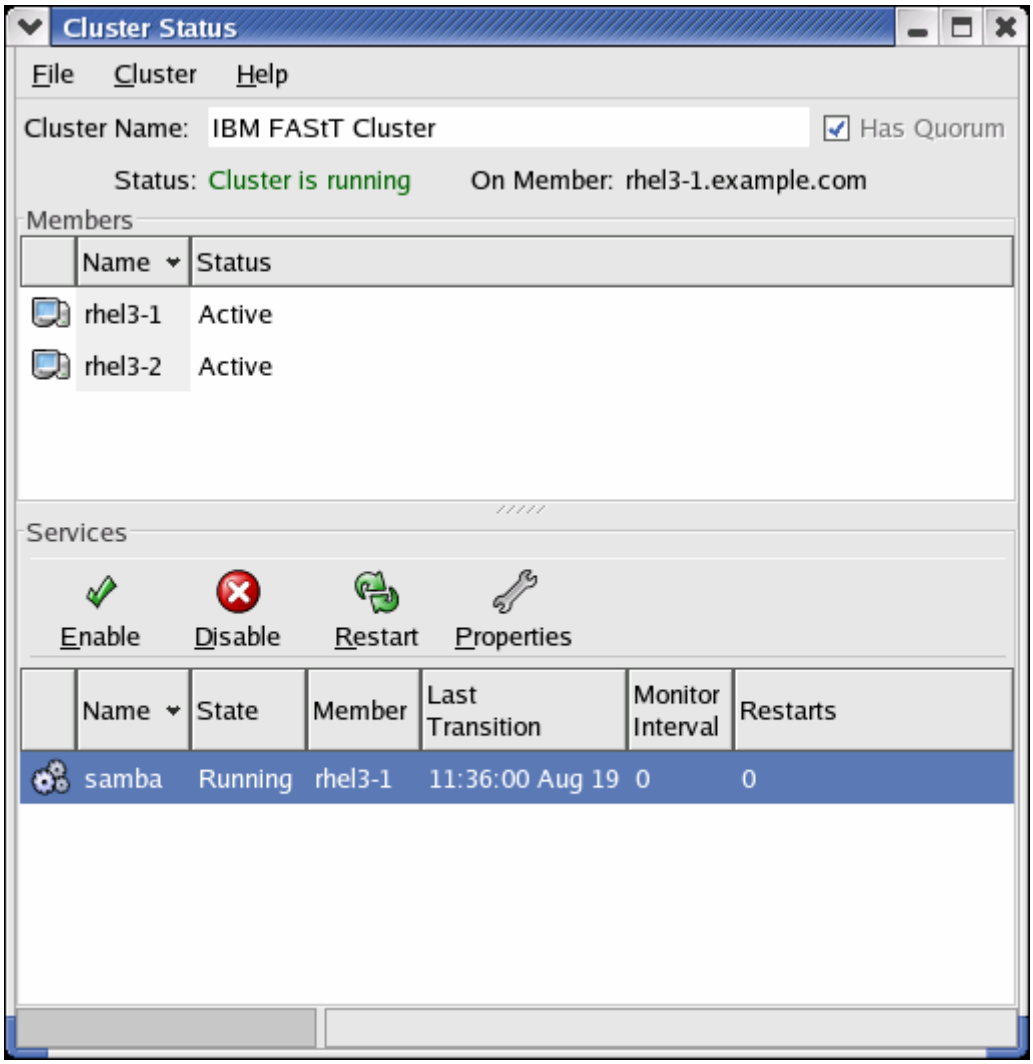

圖 25：RedHat SAMBA HA Cluster 狀態

三、測試

筆者利用另一台 Windows 的機器連線到 HA Cluster 的 service ip（圖 26），此時可以對主要的伺服器 rhel3-1 做強迫關機的動作，或是 kill clumembd 的 process 亦可模擬 rhel3-1 當機。此時 rhel3-2 便會 Take Over SAMBA 服務，達到 High Availability 目的。

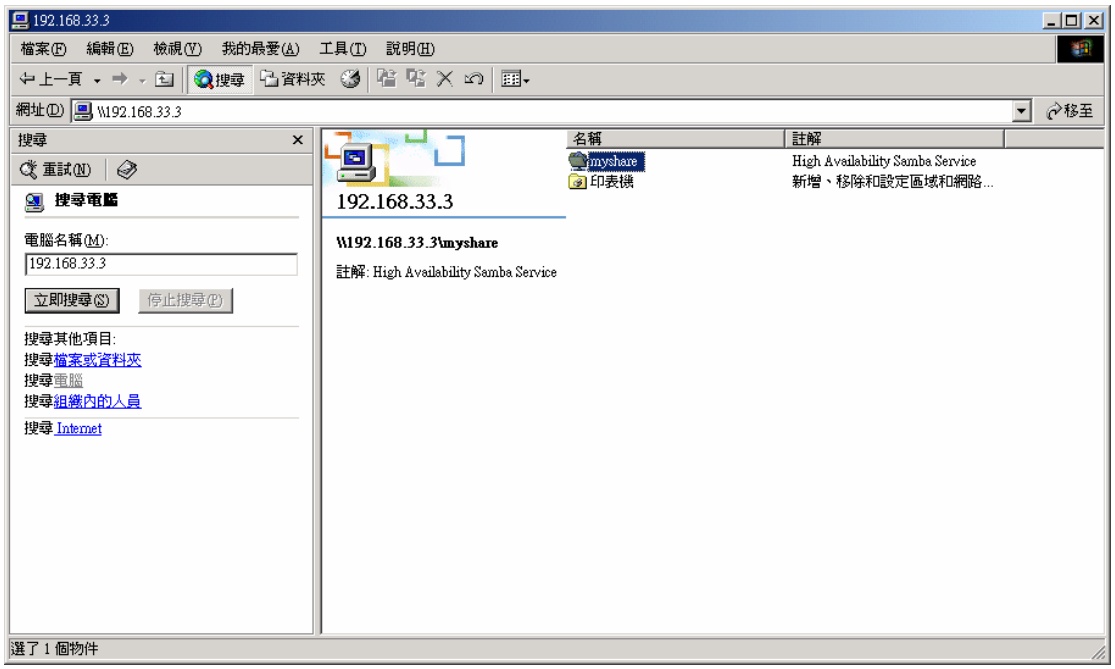

圖 26：Windows Client 連線 SAMBA HA Cluster 畫面

■ kill clumembd process

```
[root@rhel3-1 root]# ps -ef | grep clumembd
root       2597     1  0 11:46 ?        00:00:00 /usr/sbin/clumembd
root       3219  3167  0 11:48 pts/2    00:00:00 grep clumembd
[root@rhel3-1 root]# kill -9 2597
```

此時 rhel3-1 會重新開機，rhel3-2 會 take over Samba Service。我們可用「clustat」指令檢查 Cluster 狀態。

■ 檢查 Cluster 狀態

```
[root@rhel3-2 root]# clustat
Cluster Status - IBM FAStT Cluster
Cluster Quorum Incarnation #19
Shared State: Shared Raw Device Driver v1.2
Member             Status
------------------ ----------
  rhel3-1          Inactive
  rhel3-2          Active       <-- You are here

Service       Status       Owner (Last)       Last Transition      Chk Restarts
------------- -------- ---------------- ---------------------- ---- ----
samba         started        rhel3-2        11:49:31 Aug 19     0    0
```

後記

本期筆者在 SAN 架構建置 High Availability SAMBA Cluster。很多大型企業均採用 SAN 的架構儲存企業的重要資料。而 Linux 對 SAN 相關的支援亦愈來愈完整，顯示 Linu 已慢慢走入大型企業，扮演起企業的重要關鍵系統。