「Red Hat Certified Architect – Cluster 篇」

實戰講座

【iSCSI】

RHCS 中免不了會使用到儲存設備，而且 RHCS 架構中的儲存設備必須具備可讓多個主機可以同時存取的特性，筆者稱為「Share Storage」。SCSI、SSA、SAN 皆有符合此需求的相關設備,此篇文章主要介紹如何在 RHEL 4 實作 iSCSI 環境。

1 Share Storage

筆者所謂的「Share Storage」，就是符合允取讓多個主機可以同時存取的特性，不管是 SCSI、SSA…裝置，只要它可讓多台主機同時 Read/Write，便稱為「Share Storage」。

1.1 SCSI

如果 Share Disk Storage 要採用 SCSI 的裝置，必須選擇支援多主機通道（Multi-Host）。兩個 Single-initiator 的 SCSI 匯流排，在一個單一控制卡 RAID 陣列的阻絕器，一個 Single-initiator SCSI 匯流排只允許一個成員連接到它，並且提供主機的分離與比一個 Multi-initiator 匯流排具有更佳的效能表現。Single-initiator 的匯流排，確保每一個成員都能免於由於其他成員的系統負載初始或修復所引起的干擾。

圖 1 是一個無 SPOF 的 RHCS 架構， 此架構中的 Share Storage 是用兩條 single-initiator 的 SCSI 匯流排（"T"代表一個 SCSI 的 Terminator）與以確保在某台主機故障時，另一台備援的主機依舊可存取到資料。

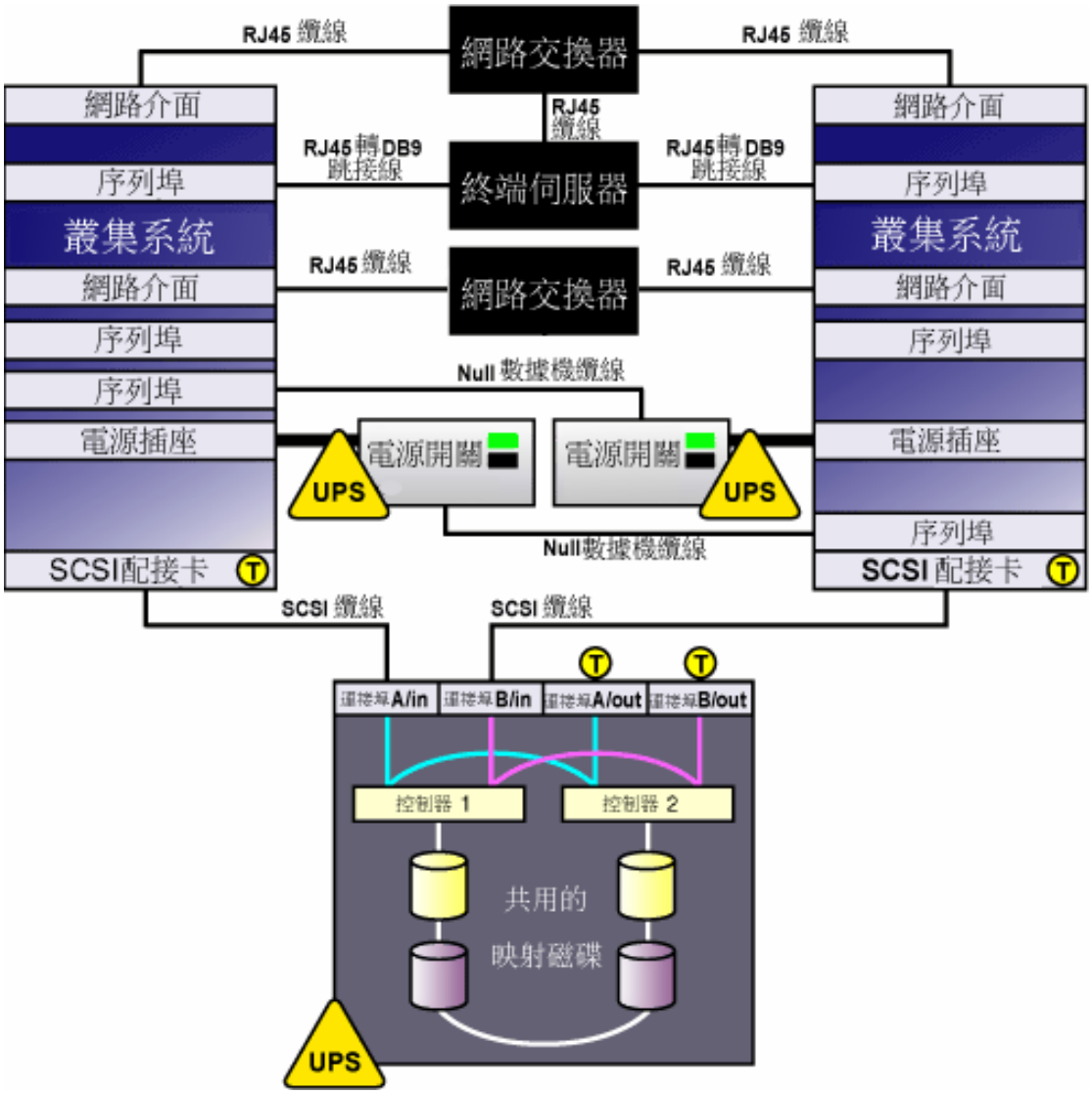

圖 1：利用 SCSI 裝置做為 Share Storage

1.2 SSA（Serial Storage Architecture）

串列式儲存架構（SSA）是由 X3 授權標準委員會的 X3T10 技術委員會中的 X3T10.1 工作群組，所發展的高效能串列式電腦與週邊介面。其最初由 IBM 研發， 現今 SSA 是由 SSA 工業協會推廣的開放技術。

SSA 基本上是一個執行於 SCSI-2 軟體協定的串列式技術。意思是 SSA 配接卡的裝置驅動程式，應該可以很容易地整合到現有的 Linux SCSI 子系統。資料是

透過以 200 MBit/s 全雙工傳輸的雙絞電纜來傳送，而這比 68 Pin 的平行 Wide SCSI 技術更易於處理。

SSA 和 SCS 相較之下，有下列優點：

- 它更易於設定和接線（它不需要終端電阻)。
- 它內建了 HA 特徵。SSA Loop 架構（相對於 SCSI 匯流排）沒有 SPOF。如果部分的 Loop 失效，裝置驅動程式將自動並透明地重新自我設定，以確保所有的 SSA 裝置可被存取而沒有任何明顯的中斷。
- 它不是使用 SCSI ID 定址，意指設定配接卡毫無困難。
- 相對於 SCSI，SSA 沒有使用匯流排裁決。而是使用類似網路的設計。資料以 128 位元組的封包被送出和接收，而且迴圈上的所有裝置可以獨立的請求 time slots。反過來 SCSI 需要匯流排裁決，如果 initiator 沒有及時釋放匯流排可能導致效能死結。
- SSA 允許每兩個裝置間相距 25 公尺。再者，允許資料穿過 50 微米的光纖電纜傳送達 2.4 公里的距離的光纖延伸器。
- 大部分的 SSA 配接卡支援兩個獨立的迴圈，使得連結鏡射的磁碟到不同迴圈以提高可用性成為可能。
- SSA 迴圈是對稱的、雙絞線自由電位的。沒有 TERMPWR 電位移的問題。

SSA 是一個比 SCSI 優良的技術，不過很可惜地，SSA 磁碟只能從 IBM 購買，取得成本太高，這些磁碟價格遠高過一般的 SCSI 磁碟價格。而且至今在 Linux 上仍沒有夠成熟的 SSA Driver。

1.3 SAN

SAN 是 Storage Area Network 的縮寫，SAN 的基礎根源於 LAN 的技術，我們今日談論的大多使用 LAN 的專業術語，如 switch、hub 和 bridge 都是現今 LAN 上使用的網路連接裝置，LAN 與 SAN 較大的差別是：LAN 對 Server 是 "Front-end" 的網路，而 SAN 對 Server 來說是 'Back-end"的網路。SAN 架構的完成必須是根基在由 ANSI（美國國家標準局）以及一些共同發展 Fibre channel 的團體對目前及未來的 Fibre channel 所制定的一些特別的規劃設計，以確保相互間的相容互動以及資料的整合。

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 是指專為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。

透過 SAN 來作 LAN-Free Backup 可以將備份的效率達到最高值，解決網路頻寬的瓶頸問題，把 LAN 的頻寬留給 Database Server 或 Mail Server 來使用；SAN Solution 規劃包含軟體及硬體，因為 SAN 產品目前在市場上並不是很相容及普及，所以在選購軟硬體時需要注意互相支援的問題；HP 在 SAN Solution 提供了 Fibre Channel Hub、Fibre Channel Host Bus Adapter、Fibre/SCSI Switch 等連接設備。

以往 MIS 人員常常為了 SCSI 的訊號不穩定及長度限制所困擾，無法滿足儲存資料量的成長速度，SAN 正好可以解決這些問題，SAN 架構在光纖通道之上，所使用的是 Fibre Channel 標準協定，一個 Loop 速度即可達 100MB/sec，長度可延伸至 30 公里，一個 Loop 可連接的裝置多達 127 個，在不用關機的狀況下即可進行硬碟陣列的儲存容量擴充。

SAN 有什麼優點 ？

- 分享資源存取與設備 - Disk and Tape。
- 高速度、距離長，可提高資料的可使用率。
- 可作 Remote Mirror 增加災難防禦力及重建速度。
- 透過 SAN 備份，降低經過 LAN 備份的 Traffic 負載。
- 集中管理與整合儲存設備資源。
- FC-Loop 可連接 127 個 Device，不需要 Shutdown Server ，即可擴充儲存容量，SAN 解決方案具備良好擴充性

下面表格可以看到 Fibre Channel 和 SCSI 的差異：

表 1：Fibre Channel 和 SCSI 比較表

Fibre Channel	SCSI
100 MB/sec per loop	20 MB/s or
200MB/sec per array	40 MB/s per array
4 wire cable structure	68 wire cable
500 meter cabling	25 meter cabling
30km with Optical Link extender	25 meter maximum
127 devices per loop	15 devices per bus
Easy HA cabling	Complex HA cabling
Non-disruptive (N.D.) expansion	N.D. not supported
No termination	Terminators required
Signal isolation	No signal isolation
Storage and networking protocol transfers	Storage only protocols
Hubs, switches, connectivity	No interconnection

2 iSCSI

企業儲存發展日新月異，早期大型伺服器的 DAS（Direct Attached Storage），為了達到儲存空間的善用及管理安裝上的效率，因而有了 SAN（Storage Area Network）的誕生，SAN 可說是 DAS 網路化發展趨勢下的產物。早先的 SAN 採用的是光纖通道（Fiber Channel；FC）技術，所以在 iSCSI 出現以前，SAN 多半單指 FC 而言。一直到 iSCSI 問市，為了方便區隔，業界才分別以 FC-SAN 及 iSCSI-SAN 的稱呼加以分辨。

iSCSI 是以 Ethernet/Internet 為實體基礎環境，以 TCP/IP 為運作協定，再往上加搭的 SCSI 資料傳輸及 SCSI 控制指令，使硬碟資源及運用達到通透於 LAN/WAN 分享的目的。

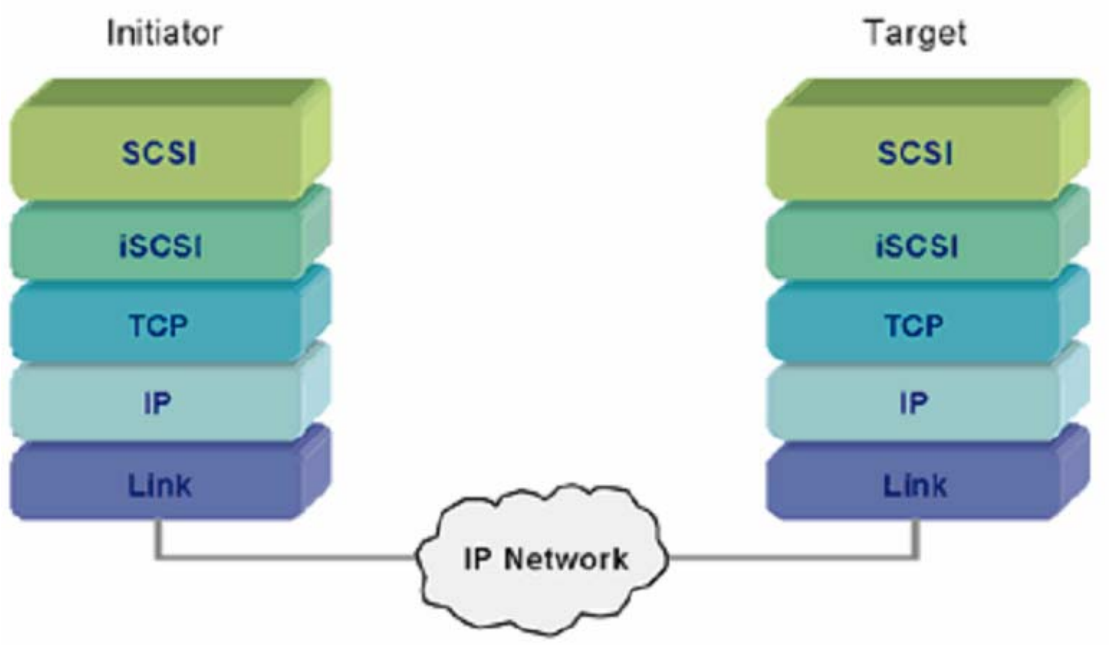

圖 2：iSCSI 架構圖

【註】iSCSI 具有 Internet SCSI、IP Storage、IP SAN、SCSI over IP、SCSI protocol over the Internet 等含意。

iSCSI 透過 IP 網路，將 SCSI 區塊資料轉換成網路封包的一種傳輸標準，它和 NAS 一樣透過 IP 網路來傳輸資料，但在資料存取方式上，則採用與 NAS 不同（File Protocol），而與 FC-SAN 相同的 Block Protocol 協定。iSCSI 最早是由

IBM 及 Cisco 於 2001 年制定的，兩家並且分別推出了支援 iSCSI 的產品－IBM IP Storage 200i 及 Cisco SN5420 Router。

雖然 IBM 在發表 iSCSI 方案產品後，也將 iSCSI 的技術規格提案交付給 IETF 審議，期望讓 iSCSI 成為 Internet 的標準，但在 2001 年至 2002 年間的審議作業階段，iSCSI 仍被人視為是 IBM 自行提出的特規專屬方案，直到 2003 年 2 月 IETF 敲定通過 iSCSI，並頒佈為 RFC 3720。iSCSI 才正式成為中立超然的網路化儲存標準。

事實上，為了解決 FC-SAN 在價格及管理上的諸多門檻，各家早有不同協定的 IP SAN 的研究開發。這些 IP SAN 的架構，其實與 iSCSI 大同小異，只不過並非走標準化的協定（事實上，在 iSCSI 標準化之前，也沒有什麼標準不標準的問題），而是各家自行研發的協定，所以基本上各家 IP SAN 是不相容的。

早先在 iSCSI 尚未標準化之前，只有少數廠商敢投注心力在 IP SAN 的開發上，但也因為每一家廠商皆開發專屬封閉協定的解決方案，所以這些方案之間並無法完全相容。而當時的市場上，由於發展 iSCSI 的廠商很少，所 以支援的平台及軟硬體等基礎設施便相當貧乏，這可說是 iSCSI 發展之初的最大阻礙及瓶頸。

據 NetApp 表示，該公司早在 2001 年年底即推出了自家的 IP SAN，它採用的是自行開發的 VLD 協定（Virtual Local Disk），儲存上屬於 Block over IP 方式。

2003 年 2 月，當 IBM 早已退出 iSCSI 之際，NetApp 宣稱推出了 iSCSI 正式標準制定之後的全球第一台 iSCSI 產品。

但接下來的兩大事件，卻被視為促進 iSCSI 大行其道的關鍵因素，那就是「**iSCSI 標準的正式通過**」，以及「**微軟的正式支援**」。

在眾所期盼的敦促下，SNIA（儲存網路產業協會；The Storage Networking Industry Associate）終於在 2003 年 2 月正式制定通過了 iSCSI 標準。而業界莫不把此標準化視為 iSCSI 發展歷程中的最關鍵因素，自此開始，有愈來 愈多的廠商開始進一步開發合乎業界標準的相關產品，iSCSI 也開始受到業界目光的青睞。

在 iSCSI 的發展過程中，除了正式標準化具 有重大意義外，微軟緊接在 Windows Server 2003 中，正式開始支援 iSCSI，並提供 iSCSI Initiator 驅動程式的下載。微軟此項深具推波助瀾的作法，帶動了整個 iSCSI 業界的發展。所以接下來，不論各類作業平台或軟硬體的支援會愈來愈齊備。

iSCSI 之所以被看好的原因不少，首先它根植於 IP 網路上，所以可以採用現有已非常成熟的管理工具及基礎建設，可為企業節省大 筆建置、管理及人事成本。更重要的是，懂 IP 的人才資源非常豐沛，成為助長 iSCSI 發展的中堅份子。此外，iSCSI 在資料傳輸距離上，幾乎沒有限制的 優點，更緊緊吸引無數企業的目光。

展望未來， iSCSI 廠商莫不期盼全新世代 10G Ethernet 的到來，因為在 10G Ethernet 的帶動下，iSCSI 的理論頻寬將會攀昇到 10Gb 的極速，那麼即使未來 FC 提昇到下世代的 4Gb，仍然不是 iSCSI 的對手。如此截然 不同的情勢逆轉，難怪讓不少廠商面露既興奮又憧憬的表情。其中，NetApp 甚至表示，未來會開始推出支援 10G 的 iSCSI 產品，此無異讓 10Gb 極速美夢成真的可能性提高不少。

3 建置 iSCSI 環境（RHEL 4）

在 SAN 中通常有兩個角色「Target」與「Initiator」：

- Target

Target 就是「儲存設備」，也就是存放資料的磁碟或是磁碟陣列，在下面利用 RHEL 4 建置 iSCSI 環境，筆者利用一台 RHEL 4 主機扮演「Target」角色。

- Initiator

Initiator 主要負責電腦主機連線到 Target 作磁碟存取功能。Initiator 可使用硬體方式 Initiator 或者軟體方式 Initiator，底下在 Linux iSCSI 實做，皆是使用軟體方式 Target 與 Initiator。

3.1 iSCSI 環境規劃

還是那句老話，先規劃！如果讀者先建置 iSCSI 環境，可先思考下列表格，先將相關資料準備好。

表 2：iSCSI 環境規劃表

功能	作業系統	主機名稱	IP Address
iSCSI-target server			
iSCSI Initiator (node1)			
iSCSI Initiator (node2)			

3.2 實作 iSCSI Target

1. 下載 iSCSI-target 軟體

首先到 http://linux-iscsi.sourceforge.net/ 網址下載 iscsitarget-0.4.5.tar.gz。

2.安裝 iSCSI-target 軟體

```
# tar zxvf iscsitarget-0.4.5.tar.gz
# cd iscsitarget-0.4.5
```

```
# export KERNELSRC=/usr/src/kernels/<kernel version>
# make && make install
```

3.建立/etc/ietd.conf

```
# vi /etc/ietd.conf
Target iqn.2008-08.com.example:storage.disk2.sys1.xyz
        Lun 0 /dev/hda# fileio
        Alias Test
```

4.在 targer 上指新增硬碟分割區（/dev/ had＃），無需格式化，/dev/had#是給 iSCSI Initiator 存取，會成為 Initiator 的一顆硬。

5.啟動 iSCSI 服務

```
# /etc/init.d/iscsi-target start.
# chkconfig iscsi-target on
```

6.檢查是否設定成功

執行「dmesg」指令後，看到下列訊息，代表設定成功。

```
#dmesg
……..
iSCSI Enterprise Target Software - version 0.4.5
iotype_init(91) register fileio
target_param(109) d 1 8192 262144 65536 2 20 8 0
```

3.3 實作 iSCSI Initiator

若是讀者有需要在 Windows 實作 iSCSI Initiator 可參考下列網址：「http://www.microsoft.com/WindowsServer2003/technologies/storage/iscsi/default.mspx」，下列筆者介紹 RHEL 4 上實作 Initiator 的方法。

1.在指 iSCSI initiator 主機上指定 initiator name

```
# echo "InitiatorName=$(iscsi-iname)" > /etc/initiatorname.iscsi
```

2.修改/etc/iscsi.conf 指定 iSCSI targer 的 IP

```
# mv /etc/iscsi.conf /etc/iscsi.conf.orig
# echo “DiscoveryAddress=<IP address of iSCSI target>” > /etc/iscsi.conf
```

3.啟動 iscsi daemon

```
# servie iscsi start
# chkconfig iscsi on
```

4.檢查

```
# cat /proc/scsi/scsi
Host: scsi1 Channel: 00 Id: 00 Lun: 00
    Vendor: LINUX      Model: ISCSI       Rev: 0
    Type: Direct-Access                   ANSI SCSI revision: 03
```

如果/proc/scsi/scsi 有出現新增 SCSI 硬碟，代表設定正確。

4 參考資料

- Red Hat Cluster Suite for Red Hat Enterprise Linux 4.5

http://www.redhat.com/docs/manuals/csgfs/browse/4.5/SAC_Cluster_Administration/index.html

- What software iSCSI target can I use with Red Hat Enterprise Linux?

http://kbase.redhat.com/faq/FAQ_85_6165.shtm

- How can I test the iSCSI software that comes with Red Hat Enterprise Linux 3?

http://kbase.redhat.com/faq/FAQ_79_6166.shtm

- What software iSCSI target can I use with Red Hat Enterprise Linux?

http://kbase.redhat.com/faq/FAQ_85_6165.shtm

- Linux 應用 iSCSI 技術（徐秉義）

http://kate.babyface.com.tw/NetAdmin/10200611iSCSI/

作者簡介

林彥明（Alex YM Lin）：現任職於 IBM，負責 HPC 超級電腦、Linux 叢集系統建置、效能調校及技術支援等工作，近來參與 NCHC IBM Cluster 1350（亞洲運算能力僅次日本的超級電腦）及中山大學 p595 HPC 超級電腦專案。具有 RHCA（Red Hat 架構師）、RHCDS（Red Hat Certified Datacenter Specialist）、RHCX（Red Hat 認證主考官）、RHCE、NCLP（Novell Linux 認證專家）、LPIC、IBM AIX …等國際認證。