



研磨 Red Hat Certified Architect—「如何大量部署 Linux」

「大量部署 Linux」！腦中出現一個問號，實務上要用在那個地方？簡單來說，只要有大量安裝 Linux 的需求，就有好好研究「如何大量部署 Linux」之必要，什麼時候會需要大量安裝 Linux？小至電腦教室，大至國家高速電腦中心的 Linux 叢集系統，或者是當今最「牛」的線上遊戲「魔獸」，都會遇到大量安裝 Linux 的需求。

很多廠商或是 Open Source 都有「大量部署 Linux」的解決方案，例如 RedHat 的「RHN」、IBM 的 CSM、Open Source 的 xCAT...等。這些軟體可輕易完成「大量部署 Linux」的工作，但俗話說「練拳不練功，到老一場空」，所以本篇文章不是要介紹這些已經被包裝過的解決方案，而是要從基礎功練起，自行打造可大量部署 Linux 的環境。



1 原理說明

「談到大量部署 Linux」，不論用那種方法，無非是希望機器開機後，連開機光碟都不用放，只需按下選項或是根本無需鍵入任何指令便自動安裝想要的作業系統。

首先第一個問題便是「不用開機光碟」，那怎麼開機？這個難題就得交給「PXE」來解決。

簡單來說，PXE (Pre-boot Execution Environment) 就是讓電腦直接透過網路 (網卡) 啟動就可開機，而無須再藉助其它開機媒體的環境。PXE 分為 Client 跟 Server 兩部份，PXE client 是燒在網卡的 ROM 中，所以第一件事必須是你的網卡有支援 PXE，現今大部份網卡都已包含 PXE Client，所以這不成問題。現在主機版大都內建網卡，只要從 BIOS 內選項中檢查，就可知道有沒有支援 Boot on LAN (圖 1)，而其所使用的協定可想成是 DHCP 跟 TFTP 的綜合體。

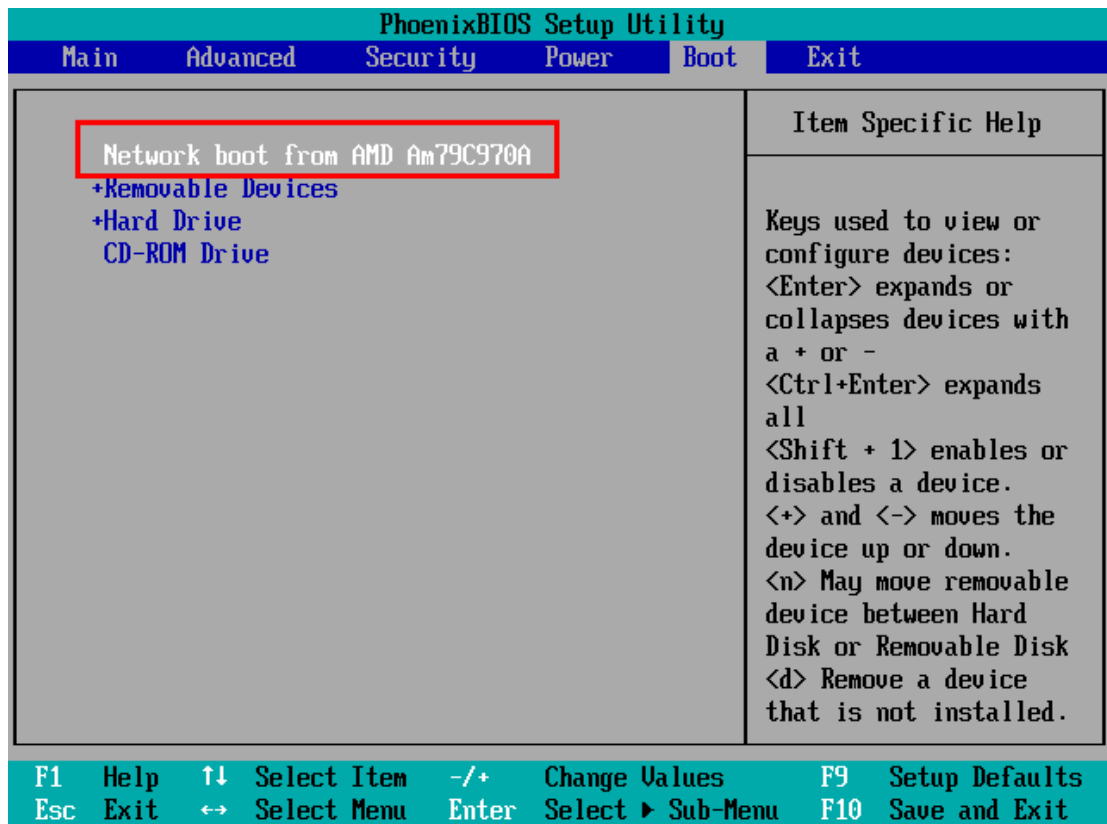


圖 1：BIOS 中 boot on LAN 的選項

PXE 與 BIOS 還有網卡之間的關聯可以由圖 2 得知，當 BIOS 把 PXE Client 載入記憶體，此時便具有 DHCP Client 及 TFTP Client 的能力。接下來便是要想法子將遠端的作業系統核心載入記憶體開機，但問題來了，既然作業系統核心是在網路上的伺服器，那麼就得透過網路下載。那麼 PXE client 如何取得 IP 位址？另外要利用那種方式下載作業系統核心？

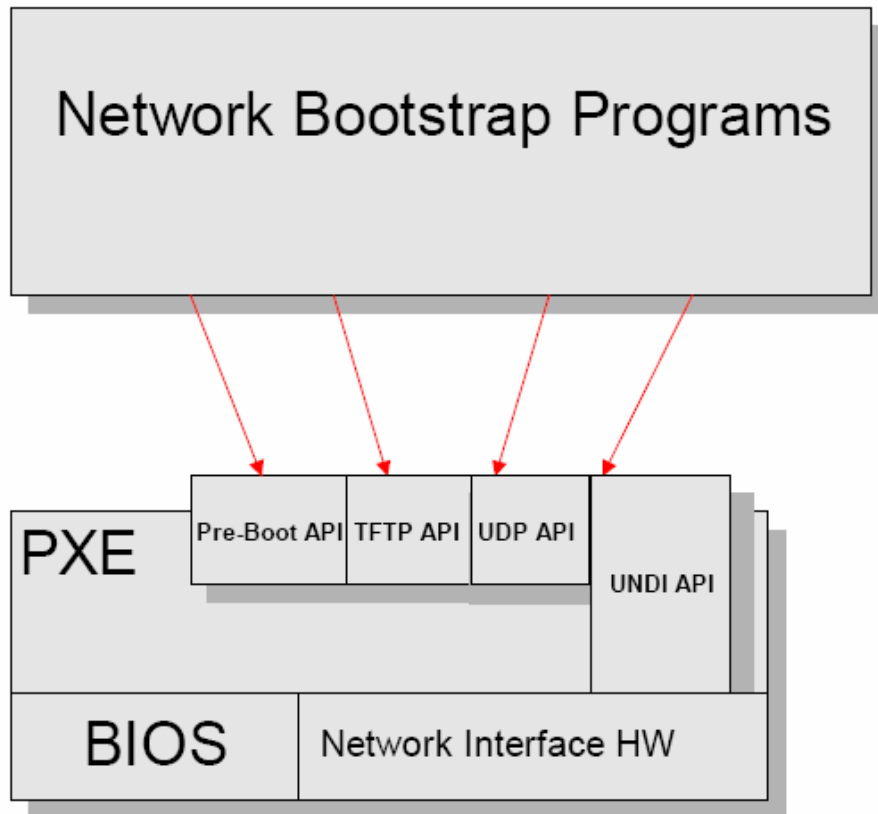


圖 2：PXE API 架構圖

圖片來源：Pre-boot Execution Environment (PXE) Specification Version 2.1

關於第一個問題，由於 PXE Client 具備 DHCP Client 能力，所以可以透過 DHCP Server 來取得 IP 位址。至於第二個問題「如何取得 kernel image」，可利用 TFTP 來取得 kernel image。整個 PXE Boot 詳細的過程，可參見圖 3。

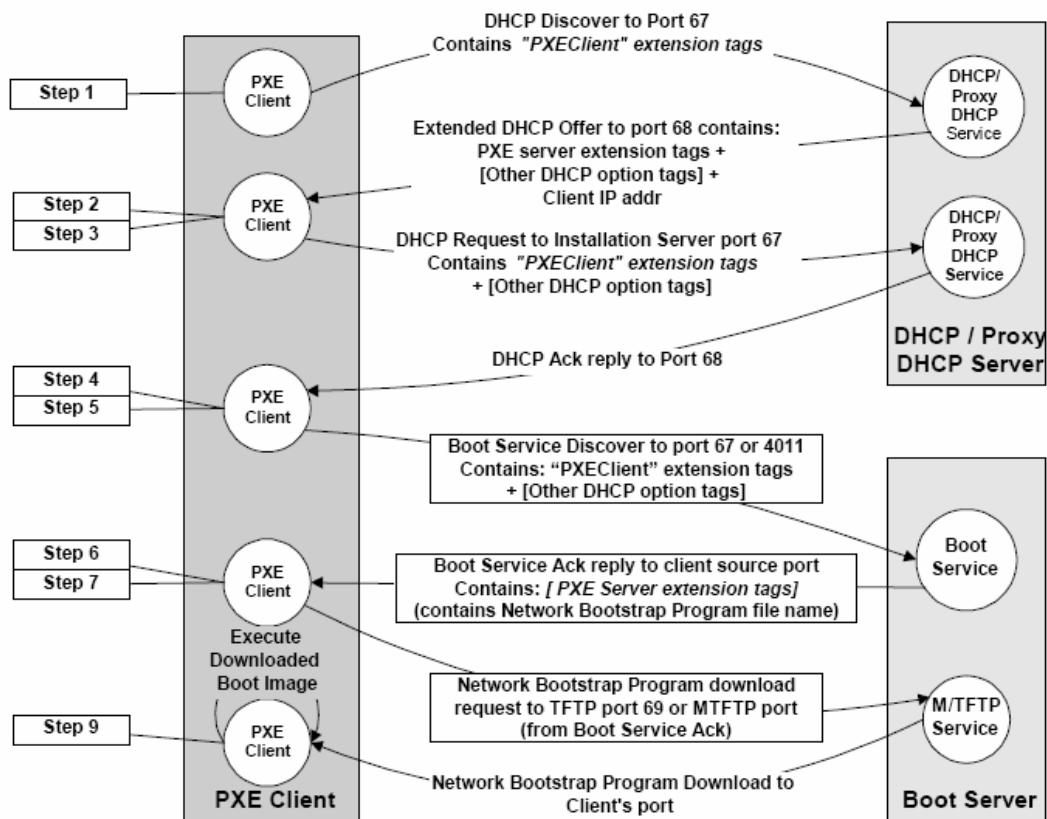


圖 3：PXE Boot 流程

圖片來源：Pre-boot Execution Environment (PXE) Specification Version 2.1

解決「無光碟開機」的問題後，接著就得煩惱如何讓它進行自動安裝，如果是「Red Hat 一族」搭配 Kickstart，便可自動安裝；如果是「SUSE 一族」，就得採用 AutoYast 的方式來自動安裝，本篇文章將會利用 Kickstart 來完成自動安裝的工作。

先來個小結，所以如果想要建置可「大量部署 Linux」的環境，考量的事項應有下列幾點：

- 主機網卡需支援 PXE Client，也就俗話說的「支援 LAN Boot」。
- 需架設 DHCP 伺服器以配發給 PXE Client IP address。
- 架設 TFTP 伺服器提供 PXE Client 開機所需的 kernel image 及相關設定檔。
- 支援 Kickstart 安裝，為了讓伺服器支援 Kickstart 安裝，則得架設 Kickstart installation Server 及編寫 Kickstart 檔案。



假設 Kickstart Installation Server 是利用 NFS 方式將 Kickstart 分享給 PXE Client，那麼整個溝通流程就如同圖 4。

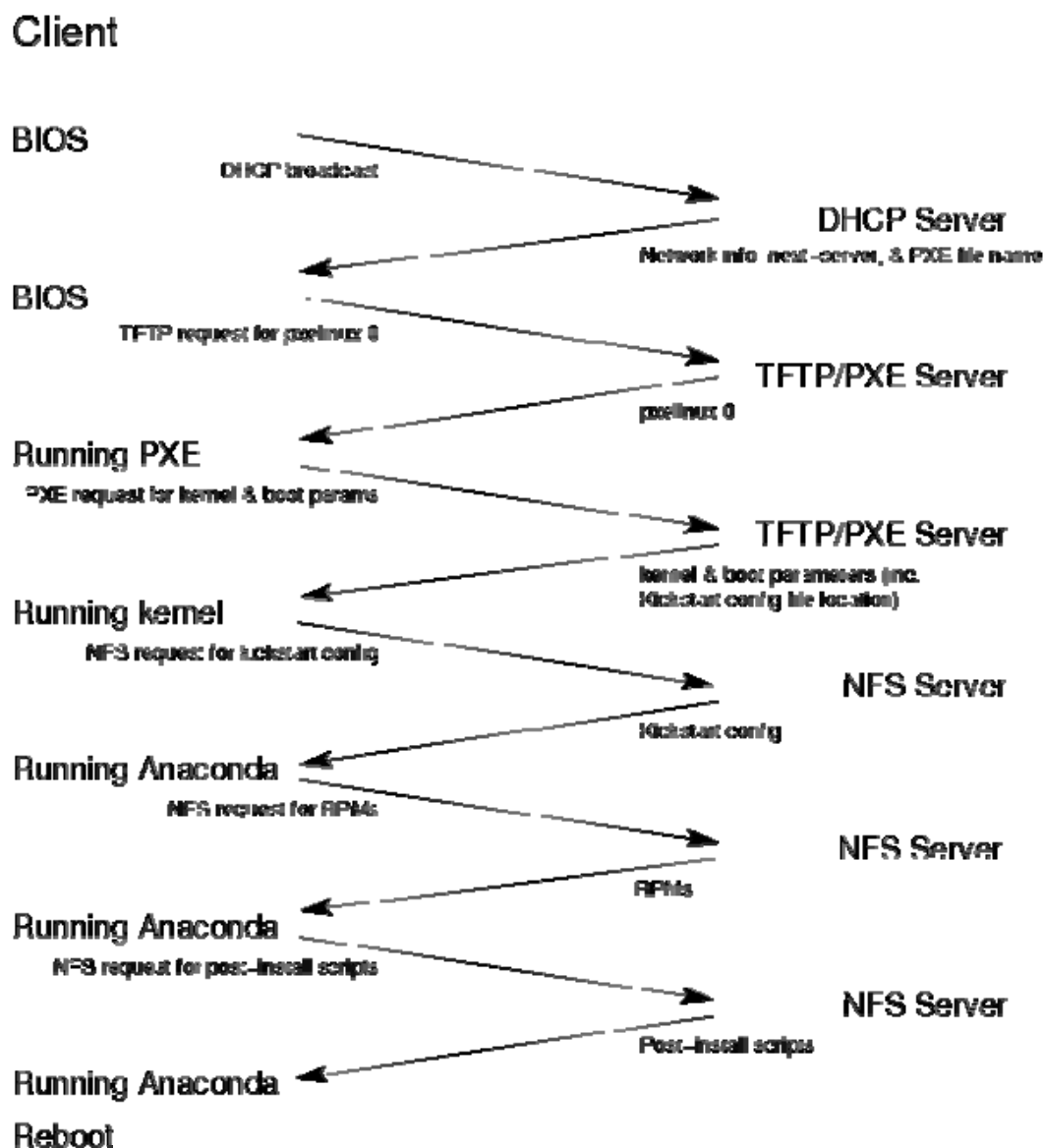


圖 4：利用 PXE 搭配 Kickstart server 自動安裝 Linux 的流程

圖片來源：<http://www.enm.bris.ac.uk/staff/pjn/Kickstart-Talk/>

圖 4 是可說是利用「PXE 來自動部署 Linux」完整示意圖，由圖 4 可知：

- 網卡得先和 DHCP Server 溝通，然後 DHCP Server 告訴 PXE Client 到哪裡（TFTP Server 的位置）去下載 pxelinux.0，pxelinux.0 是 Linux 的 boot loader，就是開機程式。



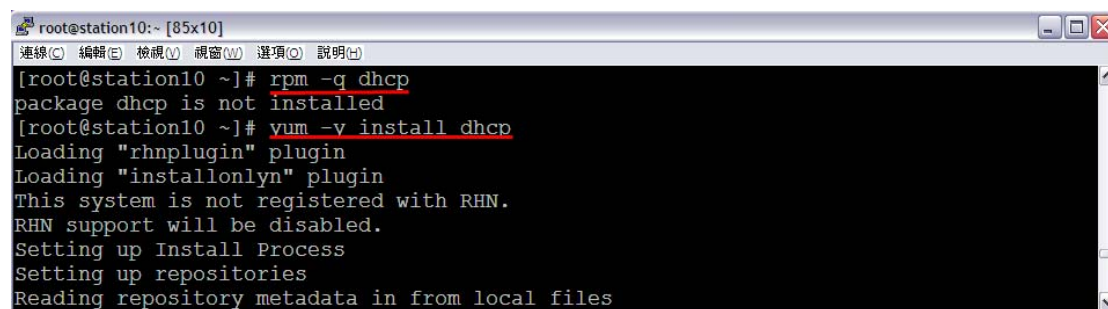
- 接著利用 TFTP 取得 pxelinux.0 檔及 Kickstart 自動安裝設定檔所在位置，把 pxelinux.0 載入記憶體，此時已如同將 Linux kernel 載入記憶體中。
- 接著 kernel 根據 DHCP Server 所告知的 Kickstart 位置，取得 Kickstart 檔案，然後根據 Kickstart 的內容來自動安裝 Linux。

下面的章節便是根據圖 4 的流程來架設及組態相關的服務 DHCP Server → TFTP Server → Kickstart Server。



2 DHCP Server 部份

利用「rpm -q dhcp」指令先確認是否已安裝 DHCP Server，若無請利用「yum -y install dhcp」安裝 dhcp Server（圖 5）。



```
root@station10:~ [85x10]
[root@station10 ~]# rpm -q dhcp
package dhcp is not installed
[root@station10 ~]# yum -y install dhcp
Loading "rhnpugin" plugin
Loading "installonlyn" plugin
This system is not registered with RHN.
RHN support will be disabled.
Setting up Install Process
Setting up repositories
Reading repository metadata in from local files
```

圖 5：安裝 dhcp 套件

```
# cp /usr/share/doc/dhcp*/dhcpd.conf.sample /etc/dhcpd.conf
```

```
# more /etc/dhcpd.conf
```

```
ddns-update-style interim;
```

```
ignore client-updates;
```

```
subnet 192.168.0.0 netmask 255.255.255.0 {
```

```
# --- default gateway
```

```
option routers 192.168.0.1;
```

```
option subnet-mask 255.255.255.0;
```

```
option nis-domain "domain.org";
```

```
option domain-name "domain.org";
```

```
option domain-name-servers 192.168.1.1;
```

```
option time-offset -18000; # Eastern Standard Time
```




```
# option ntp-servers 192.168.1.1;
# option netbios-name-servers 192.168.1.1;
# --- Selects point-to-point node (default is hybrid). Don't change this unless
# -- you understand Netbios very well
# option netbios-node-type 2;

range dynamic-bootp 192.168.0.128 192.168.0.254;
default-lease-time 21600;
max-lease-time 43200;

# we want the nameserver to appear at a fixed address
host ns {
    next-server marvin.redhat.com;
    hardware ethernet 12:34:56:78:AB:CD;
    fixed-address 207.175.42.254; }
只需在檔案最後結尾 } 前加入兩行設定
next-server 192.168.0.10; ←筆者的 TFTP Server 是 192.168.0.10
filename "pxelinux.0"; ←下載 Linux boot loader pxelinux.0
}
```

```
#service dhcpd restart
```



3 TFTP Server 部份

TFTP Server 就略嫌麻煩些，除了 pxelinux.0 這個重要的 boot loader 外，一般會希望 PXE 開機後，應該出現選單，讓使用者有所選擇，例如從原有的硬碟開機，或是利用 Kickstart 檔案重新安裝／部署這台伺服器。

pxelinux.0 是由 syslinux 套件所提供，首先檢查是否已安裝 syslinux 套件及 pxelinux.0 的存放位置。

```
[root@station10 ~]# rpm -ql syslinux-3.11-4 |grep pxe
/usr/lib/syslinux/pxelinux.0
/usr/share/doc/syslinux-3.11/pxelinux.doc
```

接著安裝 tftp-server 及 tftp 套件及啟用 tftp server

```
[root@station10 ~]# rpm -q tftp tftp-server
package tftp is not installed
package tftp-server is not installed
[root@station10 ~]# yum -y install tftp tftp-server
[root@station10 ~]# service xinetd start
[root@station10 ~]# chkconfig tftp on
```

複製開機檔案至/tftpboot 目錄下，除了將 pxelinux.0 boot loader 複製至/tftpboot，還必須將原版光碟中的/images/pxeboot 目錄中的 initrd.img 及 vmlinuz 複製到 /tftpboot 目錄。並建立/tftpboot/pxelinux.cfg 用來存放 PXE 的開機設定檔（指定 Kickstart 檔案的位置）。

```
[root@station10 ~]# ls /tftpboot/
```

initrd.img pxelinux.0 vmlinuz ←這3個檔案一定要存在

```
[root@station10 ~]# mkdir /tftpboot/pxelinux.cfg
```



建立 PXE 設定檔，PXE 設定檔預設是讀取「`/tftpboot/pxelinux.cfg/default`」來決定 PXE Menu 的選項，例如輸入「0」是從硬碟開機；輸入「1」則是重新安裝最小的作業系統；輸入「2」則是重新安裝 workstation 用途。其實為什麼可以安裝不同需求的作業系統，原理很簡單，在「`/tftpboot/pxelinux.cfg/default`」指定不同的選項（Label）對應到不同的 Kickstart 檔案。

下面便是一個實用的「`/tftpboot/pxelinux.cfg/default`」範例：

```
# cat /tftpboot/pxelinux.cfg/default
default 0 ← 預設是選項 0
prompt 1 ← 出現提示訊息
timeout 3000 ← 單位是 1/10 秒，若超過 30sec 未有任何動作，則採預設開機選項
display boot.msg ← 指定開機訊息檔為/tftpboot/boot.msg

label 0
    localboot 0

label 1
    kernel vmlinuz
    append initrd=initrd.img noipv6 ks=nfs:192.168.0.10:/var/ftp/pub/rhel5base.cfg
#192.168.0.10 為筆者的 Kickstart Server

label 2
    kernel vmlinuz
    append initrd=initrd.img noipv6 ks=nfs:192.168.0.10:/var/ftp/pub/workstation.cfg
```

PXE 訊息檔為 `/tftpboot/boot.msg`，讀者可以在 `boot.msg` 中建立 PXE Menu 的說明，讓使用者清楚知道每個選項所代表的意義。



```
# more /tftpboot/boot.msg  
  
| _\__ _|||_|_|_| |
| |)/_V_`|||V`_|_|  
| _< _/(||| _ |(|||  
| |\__\_,| |||\_,|\_|  
  
INSTALLATION MENU  
  
Choose installation type:  
  
0 Local Boot (default)  
1 RHEL5.1 Base System  
2 RHEL5.1 Workstation
```

此時，若是重開 PXE Client，則會看到下列畫面（圖 6）：

```
ip=192.168.0.251:192.168.0.10:192.168.0.1:255.255.255.0  
TFTP prefix:  
Trying to load: pxelinux.cfg/01-00-0c-29-96-44-db  
Trying to load: pxelinux.cfg/C0A800FB  
Trying to load: pxelinux.cfg/C0A800F  
Trying to load: pxelinux.cfg/C0A800  
Trying to load: pxelinux.cfg/C0A80  
Trying to load: pxelinux.cfg/C0A8  
Trying to load: pxelinux.cfg/C0A  
Trying to load: pxelinux.cfg/C0  
Trying to load: pxelinux.cfg/C  
Trying to load: pxelinux.cfg/default  
  
| |)/_V_`|||V`_|_| |
| _< _/(||| _ |(|||  
| |\__\_,| |||\_,|\_|  
  
INSTALLATION MENU  
Choose installation type:  
0 Local Boot (default)  
1 RHEL5.1 Base System  
2 RHEL5.1 Workstation  
boot: 1
```

圖 6：PXE Menu 畫面

若是此時選擇「1」，會發現畫面跟用光碟開機雷同，但最後會停在圖 7 的畫面，原因是找不到 Kickstart 檔案，最後只要再架設好 Kickstart installation server 及編寫好對應的 Kickstart 檔案便可大功告成。

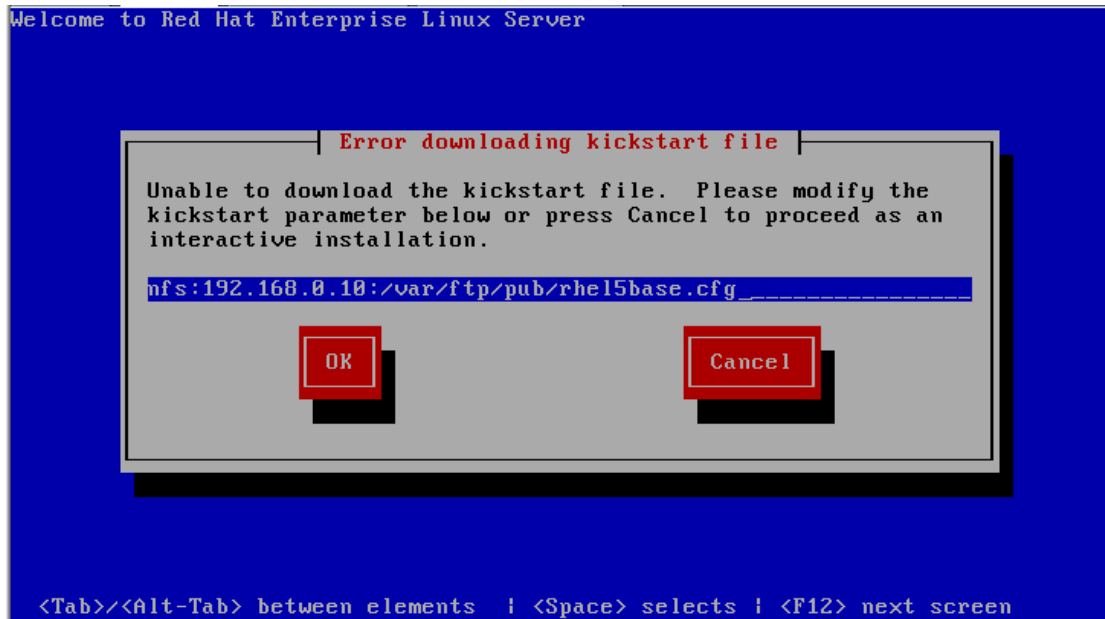


圖 7：找不到 Kickstart 檔案的畫面



4 Kickstart installation Server 部份

首先將 RedHat 安裝光碟的內容複製至 Server 上，並利用 NFS、FTP 或 HTTP 將其分享出來。

- 將 RHEL 5.X 安裝所需 RPM 全部 copy 至 Server 上

放入第 DVD 執行以下指令

```
#mount /media/dvdrom
#cp -af /mnt/dvdrom/* /var/ftp/pub/
#ln -s /var/ftp/pub /var/www/html/pub (將 /var/www/html/pub 指向 /var/ftp/pub)
```

- 利用各種方式將安裝檔案分享出來

NFS 法

```
#vi /etc/exports
/var/ftp/pub *(ro, sync) (在該檔加入此行)
#service nfs start (立即啟動 NFS Server)
#chkconfig nfs on (開機後自動啟用 NFS Server)
```

FTP 法

```
#service vsftpd start (立即啟動 FTP Server)
#chkconfig vsftpd on (開機後自動啟用 FTP Server)
```

HTTP 法

```
#service httpd start (立即啟動 HTTP Server)
#chkconfig httpd on (開機後自動啟用 HTTP Server)
```

- RHEL 5 中若是要利用 NFS 方式進行網路安裝，就必須建置 YUM Server。
筆者利用下列 script 快速產生 YUM database (repository)：

```
[root@server1 ~]# more mk_yum_server_repository.sh
#!/bin/bash
```



```
SOURCEDIR=/var/ftp/pub/
cd $ SOURCEDIR
for name in Server VT Cluster ClusterStorage
do
    cp $name/repodata/comps-rhel5-*.xml /tmp
done

# Server
cd $SOURCEDIR /Server
rm -rf repodata
createrepo -g /tmp/comps-rhel5-server-core.xml .

# VT
cd $SOURCEDIR /VT
rm -rf repodata
createrepo -g /tmp/comps-rhel5-vt.xml .

# Cluster
cd $ SOURCEDIR/Cluster
rm -rf repodata
createrepo -g /tmp/comps-rhel5-cluster.xml .

# ClusterStorage
cd $SOURCEDIR /ClusterStorage
rm -rf repodata
createrepo -g /tmp/comps-rhel5-cluster-st.xml .
```

- 最後再根據需求，在/var/ftp/pub 目錄下編寫 rhel5base.cfg 及 workstation.cfg，讓 PXE Client 根據 Kickstart 進行自動安裝。



```
#####  
# rhel5base.cfg  
# Alex YM Lin  
# 2009/07/24  
#####  
  
text  
key --skip  
keyboard us  
lang en_US  
#lang zh_TW.UTF-8  
#langsupport --default zh_TW zh_TW  
network --bootproto dhcp  
url --url ftp://192.168.0.10/pub/  
  
zerombr yes  
clearpart --all  
#part swap --size 2048  
part /boot --size 256  
part pv.01 --size=9000 --grow  
volgroup rootvg pv.01  
logvol / --vgname=rootvg --size=2048 --name=rootlv  
logvol /usr --vgname=rootvg --size=10240 --name=usrlv  
logvol /var --vgname=rootvg --size=2048 --name=varlv  
logvol /tmp --vgname=rootvg --size=1024 --name=tmpLv  
logvol /var/ftp/pub --vgname=rootvg --size=1024 --name=publv  
#logvol /home --vgname=rootvg --size=1024 --name=homelv
```




```
logvol swap --vgname=rootvg --size=2048 --name=swaplv
```

```
timezone Asia/Taipei
```

```
xconfig --resolution=1024x768 --depth=16 --startxonboot
```

```
rootpw redhat
```

```
authconfig --useshadow --enablemd5
```

```
firewall --disabled
```

```
selinux --disabled
```

```
bootloader
```

```
reboot
```

%packages

```
@ Core
```

```
openssh
```

```
openssh-server
```

```
openssh-clients
```

```
libcap
```

```
ntp
```

```
elinks
```

```
#####
```

```
# rhel5base.cfg
```

```
# Alex YM Lin
```

```
# 2009/07/24
```

```
#####
```



```
text
key --skip
keyboard us
lang en_US
#lang zh_TW.UTF-8
#langsupport --default zh_TW zh_TW
network --bootproto dhcp
url --url ftp://192.168.0.10/pub/

zerombr yes
clearpart --all
#part swap --size 2048
part /boot --size 256
part pv.01 --size=9000 --grow
volgroup rootvg pv.01
logvol / --vgname=rootvg --size=2048 --name=rootlv
logvol /usr --vgname=rootvg --size=10240 --name=usrlv
logvol /var --vgname=rootvg --size=2048 --name=varlv
logvol /tmp --vgname=rootvg --size=1024 --name=tmpLv
logvol /var/ftp/pub --vgname=rootvg --size=1024 --name=publv
#logvol /home --vgname=rootvg --size=1024 --name=homelv
logvol swap --vgname=rootvg --size=2048 --name=swaplv

timezone Asia/Taipei
xconfig --resolution=1024x768 --depth=16 --startxonboot
rootpw redhat
authconfig --useshadow --enablemd5
```



```
firewall --disabled
```

```
selinux --disabled
```

```
bootloader
```

```
reboot
```

```
%packages
```

```
*
```



5 大功告成

總算完成了！接著只要打開機器，便可根據你所輸入的選項重新自動安裝 Linux（圖 8）。雖然建置 PXE 自動安裝的環境非常辛苦，但只要建置好之後，就有很多的「利多」。想想看，假設你的伺服器有支援 IPMI 遠端電源開機的功能，根本連機房都不用進去。而且你只要修改 TFTP 的設定檔及 Kickstart 檔，很輕易就可決定被部署的 Linux 要長成什麼樣子，是要「最小安裝」還是「最大安裝」。如果一次要安裝數十台機器，在 PXE 的環境下，簡直輕而易舉！

筆者當年參與建置國網中心的 IBM Cluster 1350 時，總共得安裝數百台 Linux，當時是採用 IBM CSM 軟體來做部署的工作，其實骨子裏就是 PXE 自動安裝，那時看到一次上百台機器同時自動開機，同時開始自動安裝，著實讓人感動 ^.^。

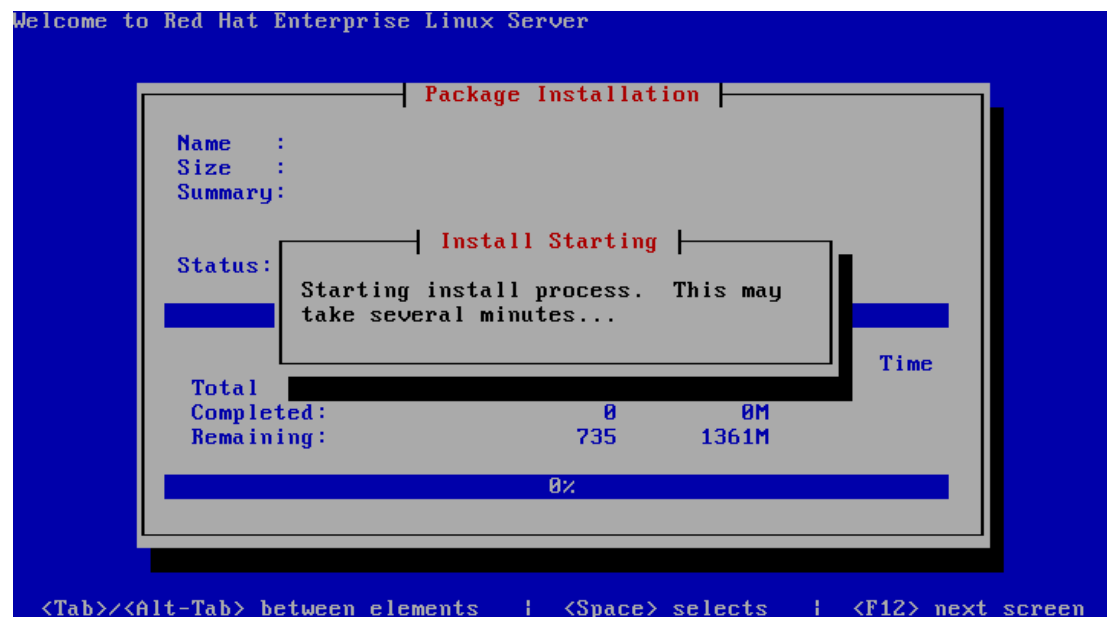


圖 8：自動安裝畫面

作者簡介

林彥明（Alex YM Lin）：IBM 高級資訊專員，負責 IBM Power System 及 HPC 超級電腦銷售支援工作，曾參與 NCHC 國網中心 IBM Cluster 1350 建置及中山



大學 p595 HPC 超級電腦專案。具有 AIX Expert、IBM MQ、RHCA (Red Hat 架構師)、RHCX (Red Hat 認證主考官) ...等國際認證。