

# xCAT 2 cookbook for pLinux on IBM Power Series

**Project: eXtreme Cluster Administration Toolkit (xCAT)**

**Document authors: Wang Xiao Peng  
Wang Hua Zhong**

**Date: 02/17/2009**

## Contents

### Table of Contents

<a href="#">1. Introduction .....</a>	<a href="#">4</a>
<a href="#">2. Install the Management node.....</a>	<a href="#">5</a>
<a href="#">2.1. Install xCAT 2 on the Management node.....</a>	<a href="#">5</a>
<a href="#">3. Setup the Management Node for node installs.....</a>	<a href="#">5</a>
<a href="#">3.1. Create a YUM repository for the RHEL node installs.....</a>	<a href="#">5</a>
<a href="#">3.1.1. Copy the iso of the Operating System which this MN running. ....</a>	<a href="#">5</a>
<a href="#">3.1.2. Mount the content of the iso to a dir.....</a>	<a href="#">5</a>
<a href="#">3.1.3. Create a yum repository file for OS image.....</a>	<a href="#">5</a>
<a href="#">3.1.4. Create a zypper repository for SLES.....</a>	<a href="#">5</a>
<a href="#">4. Setup the management node .....</a>	<a href="#">5</a>
<a href="#">4.1. Workaround the atftpd issue with p5.....</a>	<a href="#">5</a>
<a href="#">4.1.1. Remove atftp.....</a>	<a href="#">5</a>

4.1.2. Install the tftp server needed by xCAT, and restart it.....	6
4.2. Setup common attributes for xCAT in the database.....	6
4.2.1. Modify the table ppchcp (Set the default account of the HMC).....	6
4.2.2. Modify table passwd (Set the default account of the installed node).....	6
4.3. Define nodes in the Database.....	6
4.3.1. Gather Node information using the rscan command.....	7
4.3.1.1. Define HMC as an xCAT node.....	7
4.3.1.2. Discover the LPARs managed by HMC.....	7
4.3.1.3. Define xCAT node using the stanza file.....	8
4.3.2. Update the node definitions using the ctab command:.....	8
4.3.2.1. Modify the table nodelist.....	8
4.3.2.2. Modify the table nodehm.....	8
4.3.2.3. Modify the table noderes.....	8
4.3.2.4. Modify the table nodetype.....	9
4.3.2.5. Modify the table ppc and vpd.....	9
4.3.2.1. Modify the table chain.....	9
4.3.2.2. Modify the table postscripts.....	9
4.4. Setup the Management Node Services .....	9
4.4.1. Update the networks table.....	9
4.4.2. Setup the DNS .....	10
4.4.2.1. Setup /etc/hosts with entries for all you nodes, hmcs, fsps.....	10
4.4.2.2. Setup the nameserver .....	10
4.4.2.3. Setup the DNS attributes in the Site table.....	10
4.4.2.4. Setup DNS configuration .....	10
4.4.3. Configure consver.....	10
4.4.4. Check rcons.....	10
4.4.5. Update the mac table with the address of the node(s).....	11
4.4.6. Check rpower is working to the node .....	11
4.4.7. Setup dhcp.....	11
4.4.7.1. Setup the site table dhcp interfaces.....	11
4.4.7.2. Configure the DHCP.....	11
5. Install a Compute node.....	11
5.1.1. Prepare the installation source.....	12
5.1.2. Diskfull/Statefull Node installation.....	12
5.1.2.1. Set the node status.....	12
5.1.2.2. Reboot to start the installation.....	13
5.1.2.3. Check the installation results.....	13
5.1.1. Install a Diskless/Stateless Compute node.....	13

5.1.1.1. Generate the stateless image for compute node.....	13
5.1.1.2. Check the compute node packaging list.....	13
5.1.1.3. Run image generation.....	14
5.1.1.4. Edit inittab.....	14
5.1.1.5. Edit init.d/boot.rootfsck.....	14
5.1.1.6. Edit securetty.....	15
5.1.1.7. Edit fstab:.....	15
5.1.1.8. Pack the image.....	15
5.1.1.9. Set the node status.....	15
5.1.1.10. Reboot to start the installation.....	15
5.1.1.11. Check the installation result.....	15
1. Additional Commands.....	15
1.1. rflash.....	15
1.1.1. Requirements.....	15
1.1.1.1. To perform firmware update for a CEC of HMC-attached System p5 and p6	16
1.1.1.1.1. Define the CEC as a node on the management node .....	16
1.1.1.1.1. Setup SSH connection to HMC.....	16
1.1.1.1.1. Check firmware level.....	16
1.1.1.1.1. Update the firmware.....	16
1.1.1.1. To perform firmware update for a BPA of HMC-attached System p5 and p6	17
1.1.1.1.1. Define the BPA as a node on the management node. ....	17
1.1.1.1.2. Make sure that the BPA must be the “parent” attribute of one CEC which is controlled by the same HMC in PPC table. ....	18
1.1.1.1.1. Setup SSH connection to HMC.....	18
1.1.1.1.1. User rinov to check the firmware level ( see rinov manpage). ....	18
1.1.1.1.1. Update the firmware.....	18
1.1.1.1. Commit currently activated LIC update(copy T to P) for a CEC of HMC- attached System p5 and p6 .....	18
1. IB support in xCAT.....	19
1.1. IB Configuration.....	19
1.1. xDSH support IB switches.....	25
1.1. Sample Scripts.....	27
1.1.1. Annotatelog.....	27
1.1.2. getGuids .....	28

## 1. Introduction

This cookbook introduces how to use the xCAT2 to install Linux on the IBM power series machines.

The power series machine has following characters:

1. Has multiple Lpars (The Lpar will be the target machine to install, that means the Lpar will be the compute node);
2. The Ethernet card and SCSI partition are virtual devices;
3. Use the HMC or IVM as the HCP (hardware control point)

xCAT supports two types of installation type for compute node: Full installation (Statefull) and Diskless (Stateless). xCAT also supports a hierarchical cluster which uses one or more service nodes to handle the installation and management of compute nodes. **This cookbook will not cover hierarchy.**

Base on the two types of installations, the following installation scenarios will be introduced:

1. Install a stateful compute node
2. Install a stateless compute node

In order to make the reader understands the installation steps easily, this cookbook introduces a test environment as an example to simulate the whole installation process. The following is the test environment:

The management node:

Arch: Lpar of p5/p6  
OS: Red Hat 5.2  
Hostname: pmanagenode  
IP: 192.168.0.1  
HCP: HMC

The compute node:

Arch: Lpar of p5/p6  
OS: Red Hat 5.2  
Hostname: pnode1  
IP: 192.168.0.10  
HCP: HMC

The xCAT version:

xCAT-2.1+

## 2. Install the Management node

### 2.1. Install xCAT 2 on the Management node

Before preceding to setup your pLinux Cluster, you should first read [xCATtop](#) for information on downloading and installing xCAT on your Management Node.

## 3. Setup the Management Node for node installs

### 3.1. Create a YUM repository for the RHEL node installs

#### 3.1.1. Copy the iso of the Operating System which this MN running.

You can copy it from internal source server or download it from Internet

```
mkdir /iso
copy RHEL5.2-Server-20080430.0-ppc-DVD.iso to /iso/
```

#### 3.1.2. Mount the content of the iso to a dir

```
mkdir /iso/1
cd /iso
mount -o loop RHEL5.2-Server-20080430.0-ppc-DVD.iso 1
```

#### 3.1.3. Create a yum repository file for OS image

```
cd /etc/yum.repos.d
Create the /etc/yum.repos.d/rhel-Server.repo:
[rhe-5-server]
    name=RHEL 5 SERVER packages
    baseurl=file:///iso/1/Server
    enabled=1
    gpgcheck=0
```

#### 3.1.4. Create a zypper repository for SLES

Add the repository using “zypper” command:

```
zypper ar file:///iso/1 sles11

if you are creating a SLES10.2 repository
zypper sa file:///iso/1 sles10
```

## 4. Setup the management node

### 4.1. Workaround the atftpd issue with p5

The tftp client in the open firmware of p5 is only compatible with tftp-server instead of atftpd which required by xCAT2. So we have to remove the atftpd first and then install the tftp-server. This is not required for p6 or later.

#### 4.1.1. Remove atftp

```
service tftpd stop
rpm --nodeps -e atftp
```

### 4.1.2. Install the tftp server needed by xCAT, and restart it

For RHEL:

```
yum install tftp-server.ppc
```

*Notes: make sure the entry "disable=no" in the /etc/xinetd.d/tftp.*

```
service xinetd restart
```

For SLES:

```
zypper install tftp-server.ppc
```

## 4.2. Setup common attributes for xCAT in the database

### 4.2.1. Modify the table ppchcp (Set the default account of the HMC)

```
chtab hcp=hmc.cluster.net ppchcp.username=hscroot ppchcp.password=abc123
```

### 4.2.2. Modify table passwd (Set the default account of the installed node)

```
chtab key=system passwd.key=system passwd.username=root  
passwd.password=cluster
```

## 4.3. Define nodes in the Database

The definition of a node is stored in several tables of the xCAT database. There are two ways to define a node and add records in tables:

You can use the **chdef**, **mkdef** command and add them manually or use the **rscan** command to collect the data and use a stanza file.

We will use **rscan** command to get the attributes of the node and save the attributes into a stanza file. The stanza file can be used to update the database. You can then use the commands (**mkdef**, **chdef**, **lsdef**, and **rmdef**) to display or edit the information that was created in the database.

### EXAMPLE:

We want to install a node with following attributes.

**Hostname:** pnode1

**IP:** 192.168.0.10

**Arch:** ppc64 (An Lpar of power5); machine name: Server-9117-MMA-SNxxxx; Lpar

ID: 1; Lpar profile: vlpar1.prof;

**HCP:** HMC (hmc.cluster.net: 192.168.0.100)

**Install Interface:** eth0

**OS:** rhels5.2

**Postscripts:** setupntp (The scripts will be run after the installation)

**Group:** hmc,all (The node belong to these groups, then it will has the attributes of these groups as default)

**Nodetype:** compute (A common node. If you want to install a service node, the Nodetype should be "service")

The attributes of Management node:

**Hostname:** pmanagenode

**IP:** 192.168.0.1

### 4.3.1. Gather Node information using the rscan command

#### 4.3.1.1. Define HMC as an xCAT node

First, we define the hardware control point of the nodes in the cluster database.

The following command will create an xCAT node definition for an HMC with a host name of “hmc01”. The *groups*, *nodetype*, *mgt*, *username*, and *password* attributes must be set.

```
mkdef -t node -o hmc01 groups="all" nodetype=hmc mgt=hmc
      username=hscroot password=abc123
```

#### 4.3.1.2. Discover the LPARs managed by HMC

Use the **rscan** command to gather the LPAR information. This command can be used to display the LPAR information in several formats and can also write the LPAR information directly to the xCAT database. In this example we will use the “-z” option to create a stanza file that contains the information gathered by **rscan** as well as some default values that could be used for the node definitions.

To write the stanza format output of **rscan** to a file called “node.stanza” run the following command.

```
rscan -z hmc01 > node.stanza
```

This file can then be checked and modified as needed. For example you may need to add a different name for the node definition or add additional attributes and values.

***Note:** The stanza file will contain stanzas for things other than the LPARs. This information must also be defined in the xCAT database. It is not necessary to modify the non-LPAR stanzas in any way.*

The stanza file will look something like the following.

```
Server-9117-MMA-SN10F6F3D:
  objtype=node
  nodetype=fsp
  id=5
  model=9118-575
  serial=02013EB
  hcp=hmc01
  pprofile=
  parent=Server-9458-10099201WM_A
  groups=fsp,all
  mgt=hmc

pnode1:
  objtype=node
  nodetype=lpar,osi
  id=9
  hcp=hmc01
  pprofile=lpar9
  parent=Server-9117-MMA-SN10F6F3D
```

```

groups=all
mgt=hmc

pnode2:
  objtype=node
  nodetype=lpar,osi
  id=7
  hcp=hmc01
  pprofile=lpar6
  parent=Server-9117-MMA-SN10F6F3D
  groups=all
  mgt=hmc

```

***Note:** The **rscan** command supports an option to automatically create node definitions in the xCAT database. To do this the LPAR name gathered by **rscan** is used as the node name and the command sets several default values. If you use the “-w” option make sure the LPAR name you defined will be the name you want used as your node name.*

*For a node which was defined correctly before, you can use the “lsdef -z [nodename]> node.stanza” command to export the definition into the node.stanza, and use command “cat node.stanza | chdef -z” to update the node.stanza according to your need.*

#### 4.3.1.3. Define xCAT node using the stanza file

The information gathered by the **rscan** command can be used to create xCAT node definitions.

Since we have put all the node information in a stanza file we can now pass the contents of the file to the **mkdef** command to add the definitions to the database.

```
cat node.stanza | mkdef -z
```

### 4.3.2. Update the node definitions using the chtab command:

#### 4.3.2.1. Modify the table nodelist

```
chtab node="pnode1" nodelist.groups=hmc,all
```

#### 4.3.2.2. Modify the table nodehm

This lpar of power5 use the HMC as the HCP

```
chtab node="pnode1" nodehm.power=hmc nodehm.mgt=hmc nodehm.cons=hmc
```

#### 4.3.2.3. Modify the table noderes

```
chtab node="pnode1" noderes.netboot=yaboot
noderes.tftpserver=192.168.0.1 noderes.nfssserver=192.168.0.1
noderes.monserver=192.168.0.1 noderes.installnic="eth0"
noderes.primarynic="eth0" noderes.xcatmaster=192.168.0.1
```



*Note: Please make sure the attributes "installnic" and "primarynic" are set up the correct Ethernet Interface of compute node. Otherwise the compute node installation may hang requesting information from an incorrect interface.*

#### 4.3.2.4. Modify the table nodetype

```
chtab node="pnode1" nodetype.os="rhels5.2" nodetype.arch="ppc64"  
      nodetype.profile="compute" nodetype.nodetype="lpar,osi"
```

*Note: This means when node: pnode1 is installed, it will install the rhels5.2 OS and architecture: ppc64.*

#### 4.3.2.5. Modify the table ppc and vpd

##### How to obtain the node attributes from the HMC

To achieve the ARCH attributes in example, we list all the lpar attributes of the node. You can follow these steps to obtain the lpar attributes of a new Lpar.

1. Login the HMC
2. Use the "ssh HMC-l hscroot id to login" .
3. Display the machines managed by this HMC  
lssyscfg -r sys
4. Display the lpars of the machine  
lssyscfg -r lpar -m Server-9117-MMA-SNxxxx

##### Modify the table ppc

```
chtab node="pnode1" ppc.hcp=hmc.cluster.net ppc.id=1  
      ppc.pprofile=vlpar1.prof ppc.parent=Server-9117-MMA-SNxxxx
```

##### Modify the table vpd

```
chtab node=Server-9117-MMA-SNxxxx vpd.serial=xxxx vpd.mtm=9117-MMA
```

Note: the mtm and serial num come from the output of 5.1.5.1

#### 4.3.2.1. Modify the table chain

```
chtab node="pnode1" chain.node="pnode1" chain.currstate=boot  
      chain.currchain=boot
```

#### 4.3.2.2. Modify the table postscripts

```
chtab node=pnode1 postscripts.postscripts=setupntp
```

This only needs to be done, if you want NTP setup on the compute node.

## 4.4. Setup the Management Node Services

### 4.4.1. Update the networks table

EXAMPLE:

An interface which needs to be added:

```
eth1    Link encap:Ethernet HWaddr FE:99:72:0C:8B:04  
        inet addr:192.168.0.1 Bcast:192.168.0.255 Mask:255.255.255.0  
        inet6 addr: fe80::fc99:72ff:fe0c:8b04/64 Scope:Link  
        UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1  
        RX packets:5373584 errors:0 dropped:0 overruns:0 frame:0
```

```
TX packets:10583411 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:406081967 (387.2 MiB) TX bytes:15279683348 (14.2 GiB)
Interrupt:20
```

You can use the `chtab` or `tabedit` command to add the network entry:

```
chtab net=192.168.0.0 networks.netname=net1 networks.mask=255.255.255.0
networks.mgtifname=eth1 networks.gateway=192.168.0.1
networks.dhcpserver=192.168.0.1 networks.tftpserver=192.168.0.1
networks.nameservers=192.168.0.1
```

## 4.4.2. Setup the DNS

### 4.4.2.1. Setup /etc/hosts with entries for all you nodes, hmcs, fsps

```
127.0.0.1 localhost
192.168.0.1 pmanagenode
192.168.0.10 pnode1
```

### 4.4.2.2. Setup the nameserver

Add following lines into `/etc/resolv.conf`

```
search cluster.net
nameserver 192.168.0.1
```

### 4.4.2.3. Setup the DNS attributes in the Site table

```
chtab key=nameservers site.value=192.168.0.1 (Setup local machine as
nameserver)
chtab key=forwarders site.value=9.114.1.1 (Setup the external
nameserver)
chtab key=domain site.value=cluster.net (Setup the local domain)
```

### 4.4.2.4. Setup DNS configuration

```
makedns
service named start
chkconfig --level 345 named on
```

## 4.4.3. Configure consverver

```
makeconservercf
service consverver restart
```

## 4.4.4. Check rcons

```
rcons pnode1
```

*If the output is:*

Can't locate object method "stty" via package "IO::Tty" at  
/opt/xcat/lib/perl/xcAT/PPCcli.pm line xxx.

*Download the package “perl-IO-Stty-.02-1.2.el5.rf.noarch.rpm” and install the package manually, following the step:*

```
rpm -ivh perl-IO-Stty-.02-1.2.el5.rf.noarch.rpm
```

#### 4.4.5. Update the mac table with the address of the node(s)

If there's only one ethernet adapter on the node:

```
getmacs pnode1
```

But, if there're more than one ethernet adapters on the node, you have to specify more parameters like this:

```
getmacs pnode1 -S 192.168.0.1 -G 192.168.0.1 -C 192.168.0.10
```

The output looks like following:

```
pnode1:
```

Type	Location	Code	MAC Address	Full Path Name	Ping Result
Device	Type				

ent U9133.55A.10E093F-V4-C5-T1	f2:60:f0:00:40:05	/vdevice/1-lan@30000005	virtual		
--------------------------------	-------------------	-------------------------	---------	--	--

#### 4.4.6. Check rpower is working to the node

```
rpower pnode1 stat
```

```
rpower pnode1 on/off
```

#### 4.4.7. Setup dhcp

##### 4.4.7.1. Setup the site table dhcp interfaces

```
chtab key=dhcpinterfaces site.value='pmanagenode|eth1'
```

##### 4.4.7.2. Configure the DHCP

Add the relevant networks into the DHCP configuration:

```
makedhcp -n
```

Add the defined node into the DHCP configuration:

```
makedhcp -a
```

```
service dhcpd restart
```

*Note: Please make sure there is only one dhcpd server running in the subnet*

## 5. Install a Compute node

### EXAMPLE:

**Hostname:** pnode1

**IP:** 192.168.0.10

**Arch:** ppc64 (An Lpar of power5); machine name: Server-9117-MMA-SNxxxx; Lpar ID: 1; Lpar profile: vlp1.prof;

**HCP:** HMC (hmc.cluster.net: 192.168.0.100)

**Install Interface:** eth0

**OS:** rhels5.2

**Postscripts:** setupntp (The scripts will be run after the installation)

**Group:** hmc,all (The node belong to these groups, then it will has the attributes of these groups as default)

**Nodetype:** compute (A common node. If you want to install a service node, the Nodetype should be “service”)

The attributes of Management node:

**Hostname:** pmanagenode

**IP:** 192.168.0.1

An example for pnode1.stanza

pnode1:

```
objtype=node
arch=ppc64
cons=hmc
power=hmc
currchain=boot
currstate=boot
groups=hmc,all
hcp=hmc.cluster.net
id=1
mgt=hmc
nodetype=lpar,osi
os=rhels5.2
parent=Server-9117-MMA-SNxxxx
mtm=9117-MMA
serial=xxxx
power=hmc
pprofile= vlp1.par1.prof
profile=compute.ppc64
installnic=eth0
primarynic=eth0
postscripts=setupntp
netboot=yaboot
xcatmaster=192.168.0.1
tftpserver=192.168.0.1
monserver=192.168.0.1
nfssserver=192.168.0.1
```

### 5.1.1. Prepare the installation source

```
copycds /iso/RHEL5.2-Server-20080430.0-ppc-DVD.iso
```

*Note: If you encounter the issue that the iso cannot be mounted by the copycds command. Make sure the SELinux is disabled. See .*

### 5.1.2. Diskfull/Statefull Node installation

#### 5.1.2.1. Set the node status

```
nodeset "pnode1" install
```

#### 5.1.2.2. Reboot to start the installation

```
rpower "pnode1" boot
```

#### 5.1.2.3. Check the installation results

1. SSH service on the node is working and you can login without a password
2. If ssh is not working, force exchange the ssh key to the compute node using xdsh:

```
xdsh pnode1 -K
```

Note: If you cannot ssh into the node without a password, then when you run this command to exchange the keys. At that point you should be able to run

```
xdsh pnode1 date
```

### 5.1.1. Install a Diskless/Stateless Compute node

#### 5.1.1.1. Generate the stateless image for compute node

We will build the stateless image on the Management Node. It will have the same OS and architecture of the Management node. If you need another OS image or architecture than is installed on the Management Node, you will need a machine that meets the architecture you want for the image and create the image on that node.

#### 5.1.1.2. Check the compute node packaging list

```
cd /opt/xcat/share/xcat/netboot/rh/
```

For SLES:

```
cd /opt/xcat/share/xcat/netboot/sles/
```

Check to make sure compute.exlist excludes the packages you want to exclude.

Check to make sure compute.pkglist has the following packages to install

```
bash
nfs-utils
stunnel
dhclient
kernel
openssh-server
openssh-clients
busybox-anaconda
wget
vim-minimal
ntp
```

And add any other packages you would like to install on your compute node. For example, if you want to have userids with passwords you should add the following:

```
cracklib
```

```
libuser
passwd
```

For SLES11, make sure the files `compute.sles11.ppc64.pkglist` and `compute.sles11.ppc64.otherpkgs.pkglist` exist. And make sure `compute.sles11.ppc64.pkglist` has the following packages to install

```
aaa_base
bash
nfs-utils
dhcpcd
kernel-ppc64
openssh
psmisc
wget
sysconfig
syslog-ng
klogd
vim
```

Make sure `compute.sles11.ppc64.otherpkgs.pkglist` has at least the following packages to install

```
stunnel
```

#### 5.1.1.3. Run image generation

```
cd /opt/xcat/share/xcat/netboot/rh
./genimage -i eth0 -n ibmveth -o rhels5.2 -p compute
```

For SLES11:

```
./genimage -i eth0 -n ibmveth -o sles11 -p compute
```

For SLES11, the rpm package of `stunnel` should be put into the directory `/install/post/otherpkgs/sles11/ppc64`. Since SLES11 doesn't ship `stunnel` with it, you have to build the rpm package from the source code. You can get the source code of `stunnel` from <http://www.stunnel.org/download/source.html>.

#### 5.1.1.4. Edit inittab

This is only for SLES11, ignore it if the diskless image is for Red hat.

```
cd /install/netboot/sles11/ppc64/compute/rootimg/etc
```

Uncomment the console line in the file `inittab`:

```
cons:12345:respawn:/sbin/smart_agetty -L 38400 console
```

#### 5.1.1.5. Edit init.d/boot.rootfsck

This is only for SLES11, ignore it if the diskless image is for Red hat.

```
cd /install/netboot/sles11/ppc64/compute/rootimg/etc
```

Edit the file `init.d/boot.rootfsck` if it exists, change `"MAY_FSCK=1"` to `"MAY_FSCK=0"` in the line 40.

#### 5.1.1.6. Edit securetty

This is only for SLES11, ignore it if the diskless image is for Red hat.

Add “console” entry to the end of the file securetty if there's no “console” entry in the file.

#### 5.1.1.7. Edit fstab:

`cd /install/netboot/rhels5.2/ppc64/compute/rootimg/etc/`

Add following lines into the fstab:

```
proc /proc proc rw 0 0
sysfs /sys sysfs rw 0 0
devpts /dev/pts devpts rw,gid=5,mode=620 0 0
compute_ppc64 / tmpfs rw 0 1
```

#### 5.1.1.8. Pack the image

`packimage -o rhels5.2 -p compute -a ppc64`

or for SLES

`packimage -o sles11 -p compute -a ppc64`

#### 5.1.1.9. Set the node status

`nodeset "pnode1" netboot`

#### 5.1.1.10.Reboot to start the installation

`rpower "pnode1" boot`

#### 5.1.1.11.Check the installation result

SSH service to the compute node is working. You can ssh login without a password.

## 1. Additional Commands

### 1.1. rflash

rflash - Performs Licensed Internal Code (LIC) update support for HMC-attached POWER5 and POWER6 Systems.

#### 1.1.1. Requirements

POWER5 and POWER6 Licensed Internal Code updates must meet the following prerequisites:

1. Enable the HMC to allow remote ssh connections.
2. Ensure that ssh is installed on the AIX xCAT management node. If you are using an AIX management node, make sure the value of "useSSHonAIX" is “yes” in the site table.

**chtab key="useSSHonAIX" site.value=yes**

3. The Lpar , CEC, or BPA has been defined in the **nodelist, nodehm,nodetype,vpd, ppc** tables

4. Define the HMC related the above node as a node on the management node. For example,

**nodeadd** hmc01.clusters.com groups=hmc

#### 5. Setup SSH connection to HMC

Run the **rspconfig** command to set up and generate the ssh keys on the xCAT management node and transfer the public key to the HMC. You must also manually configure the HMC to allow remote ssh connections. For example:

**rspconfig hmc01.clusters.com sshcfg=enable**

#### 6. Get the Microcode update package and associated XML file.

### 1.1.1. To perform firmware update for a CEC of HMC-attached System p5 and p6

#### 1.1.1.1. Define the CEC as a node on the management node .

**Update the xCAT required xCAT tables:**

**Modify the nodelist table**

**nodeadd** Server-m\_tmp-SNs\_tmp groups=hmc,all

**Modify the table nodehm**

**chtab** node="Server-m\_tmp-SNs\_tmp" nodehm.mgt="hmc"

**Modify the table nodetype:**

**chtab** node="Server-m\_tmp-SNs\_tmp" nodetype.nodetype="fsp"

**Modify the table ppc:**

**chtab** node="Server-m\_tmp-SNs\_tmp" ppc.hcp= hmc01.clusters.com

**Modify the tab vpd:**

**chtab** node=Server-m\_tmp-SNs\_tmp vpd.serial=s\_tmp vpd.mtm=m\_tmp

**Set the account of the HMC(Modify the ppchcp):**

**chtab** hcp=hmc01.clusters.com ppchcp.username=hscroot ppchcp.password=abc123

#### 1.1.1.1. Setup SSH connection to HMC

Generate the ssh keys on the xCAT management node and transfer the public key to the HMC to configure the HMC to allow remote ssh connections. See Setup SSH connection to HMC.

#### 1.1.1.1. Check firmware level

**rinv** Server-m\_tmp-SNs\_tmp firm

#### 1.1.1.1. Update the firmware



Download the Microcode update package and associated XML file from the IBM Web site: <http://www14.software.ibm.com/webapp/set2/firmware/gjsn>. Create the **/tmp/fw** directory, if necessary, and copy the downloaded files to the **/tmp/fw** directory.

Run the **rflash** command with the **--activate** flag to specify the update mode to perform the updates. ( Please see the “**rflash**” manpage for more information )

```
rflash Server-m_tmp-SNs_tmp -p /tmp/fw --activate disruptive
```

**NOTE:**You Need check your update is concurrent or disruptive here!! other commands sample:

```
rflash Server-m_tmp-SNs_tmp -p /tmp/fw --activate concurrent
```

**Notes:**

- 1) If the noderange is the group lpar, the upgrade steps are the same as the CEC's.
- 2) System p5 and p6 updates can require time to complete and there is no visual indication that the command is proceeding.

### **1.1.1. To perform firmware update for a BPA of HMC-attached System p5 and p6**

#### **1.1.1.1. Define the BPA as a node on the management node.**

**Update the xCAT tables:**

**Modify the nodelist table. Define the BPA as a node**

```
nodeadd Server-m_tmps_tmp groups=hmc,all
```

**Modify the table nodehm**

```
chtab node="Server-m_tmps_tmp" nodehm.mgt="hmc"
```

**Modify the table nodetype:**

```
chtab node="Server-m_tmps_tmp" nodetype.nodetype="fsp"
```

**Modify the table ppc:**

```
chtab node="Server-m_tmps_tmp" ppc.hcp= hmc01.clusters.com ppc.id=x
```

**Modify the tab vpd:**

```
chtab node=Server-m_tmps_tmp vpd.serial=s_tmp vpd.mtm=m_tmp
```

**Set the account of the HMC(Modify the ppchcp):**

```
chtab hcp=hmc01.clusters.com ppchcp.username=hscroot ppchcp.password=abc123
```

**1.1.1.2. Make sure that the BPA must be the “parent” attribute of one CEC which is controlled by the same HMC in PPC table.**

If not, add an item in the table ppc:

```
chtab node="Server-mtmp-SNstmp" ppc.hcp=
hmc01.clusters.com ppc.id=n ppc.parent="Server-
m_tmtps_tmp"
```

**1.1.1.1. Setup SSH connection to HMC**

Generate the ssh keys on the xCAT management node and transfer the public key to the HMC to configure the HMC to allow remote ssh connections.

See Setup SSH connection to HMC.

**1.1.1.1. User rinv to check the firmware level ( see rinv manpage).**

```
rinv Server-m_tmtps_tmp firm
```

**1.1.1.1. Update the firmware**

**Download the Microcode update package and associated XML file from the IBM Web site:**

<http://www14.software.ibm.com/webapp/set2/firmware/gjsn>.

Create the **/tmp/fw** directory, if necessary, and copy the downloaded files to the **/tmp/fw** directory.

**Run the rflash command with the --activate flag to specify the update mode to perform the updates.**

```
rflash Server-m_tmtps_tmp -p /tmp/fw --activate disruptive
```

**NOTE:**You Need check your update is concurrent or disruptive here!! other commands sample:

```
rflash Server-m_tmtps_tmp -p /tmp/fw --activate concurrent
```

**1.1.1. Commit currently activated LIC update(copy T to P) for a CEC of HMC-attached System p5 and p6**

1. See To perform firmware update for a CEC of HMC-attached System p5 and p6 steps ( 1 – 4).

2. Check the output of the last step, to check whether the LIC will be committed. If yes the run the rflash command with the `--commit` flag.

```
rflash Server-m_tmp-SNs_tmp --commit
```

**Notes:**

- (1) If the noderange is BPA/Lpar, the commit steps are the same as the CEC's.
- (2) If to recover the installed LIC updates, the 1-6 steps of 9.1.4. Only the last step is different.
- (3) At present, the command “**licutil**” on HMC doesn't support for the “**commit**” and “**recover**” operation on the BPA. And “**rflash**” is dependent on the “**licutil**” command. So When the `--commit` or `--recover` two flags is used, the **noderange** cannot be a BPA or BPA list in the “**rflash**” command. xCAT will pay the attention to the changes all the time.

## 1. IB support in xCAT

### 1.1. IB Configuration

We will use one sample postscript configiba to config IB secondary adapter. This script can run on AIX and Linux managed nodes both.

This sample script configiba is stored in `/opt/xcat/share/xcat/ib/scripts`, so user needs to manually copy sample script configiba to `/install/postscript` first. The method to configure IB adapters is to either during node installation or trigger command `updatenode` after node installation which will finally trigger configiba also.

To use this script, user need to firstly setup DNS for the new adapters before triggering command `updatenode`:

1. The IP address entries for IB interfaces in `/etc/hosts` on xCAT managed nodes should have the node short hostname and the unique IB interface name in them. The format should be `<ip_address_for_this_ib_interface node_short_hostname-ib_interfacename>`.

For example:

```
c890f11ec01 is the node short hostname, c890f11ec01-ib0, c890f11ec01-ib1, c890f11ec01-ib2, etc. are the IP names for the IB interfaces on c890f11ec01.
```

## 2.Update networks table with IB sub-network

For example:

```
chtab net=172.16.0.0 networks.netname=ib0 networks.mask=255.255.0.0
networks.mgtifname=ib
```

*Note: Attributes gateway, dhcpserver, tftpserver, and nameservers in networks table are not a MUST to be assigned, since the xCAT management work is still running on ethernet.*

## 3.On AIX, change the default connection between management nodes and computer nodes from ssh to rsh:

```
chtab key=useSSHonAIX site.key=no
```

4.If the computer node have already installed and up and running, make sure /etc/resolv.conf is available on the compute node before running updatenode, since configiba script will connect to name server to resolve IP address for the IB interfaces. If not, define /etc/resolv.conf on compute node or use rcp to copy resolv.conf from management node as a workarround like following:

```
domain ppd.pok.ibm.com
search ppd.pok.ibm.com
nameserver 172.16.0.1
```

*Note: 172.16.0.1 is the name server address which provide the IP addresses for IB interfaces on compute nodes.*

## 5.Add the entries in the /etc/hosts into DNS and restart the DNS

Following is an example of /etc/hosts

```
192.168.0.10      c890f11ec01-ib0
192.168.0.11      c890f11ec01-ib1
```

For Linux Managed Nodes:

```
makedns
service named restart
```

For AIX Managed Nodes:

```
makedns
stopsrc -s named
startsrc -s named
lssrc -s named
```

***Note: Only if the state of named is active, we can continue to test DNS.***

## 6.Check if DNS for the IB network has been setup successfully

```
nslookup c890f11ec01-ib0
nslookup c890f11ec01-ib1
```

7. For RHEL and SLES, prepare the IB drivers/libraries for compute nodes. The packages have been listed in the below table.

a. put IB drivers/libraries rpms under /install/post/otherpkgs/<os>/<arch> directory where <os> and <arch> can be found in the nodetype table.

b. add rpm names (without version number) into /install/custom/install/<ostype>/profile.otherpkgs.pkglist, where <profile> is defined in the nodetype table. <ostype> is the operating system name without the version number. The following os types are recognized by xCAT.

```
centos
fedora
rh
sles
windows
```

IB drivers/libraries list:

On RHEL:

Driver/Library	Corresponding rpms in RHEL5.3	
<b>openib</b>	<i>openib-*.el5.noarch.rpm</i>	
<b>libib</b>	<b>32bit</b>	<i>libibcm-*.el5.ppc.rpm</i>
		<i>libibcm-devel-*.el5.ppc.rpm</i>
		<i>libibcm-static-*.el5.ppc.rpm</i>
		<i>libibcommon-*.el5.ppc.rpm</i>
		<i>libibcommon-devel-*.el5.ppc.rpm</i>
		<i>libibcommon-static-*.el5.ppc.rpm</i>
		<i>libibmad-*.el5.ppc.rpm</i>
		<i>libibmad-devel-*.el5.ppc.rpm</i>
		<i>libibmad-static-*.el5.ppc.rpm</i>
		<i>libibumad-*.el5.ppc.rpm</i>
		<i>libibumad-devel-*.el5.ppc.rpm</i>
		<i>libibumad-static-*.el5.ppc.rpm</i>
		<i>libibverbs-*.el5.ppc.rpm</i>
		<i>libibverbs-devel-*.el5.ppc.rpm</i>
		<i>libibverbs-static-*.el5.ppc.rpm</i>
		<i>libibverbs-utils-*.el5.ppc.rpm</i>
	<b>64bit</b>	<i>libibcm-*.el5.ppc64.rpm</i>
		<i>libibcm-devel-*.el5.ppc64.rpm</i>

		<i>libibcm-static-*.el5.ppc64.rpm</i> <i>libibcommon-*.el5.ppc64.rpm</i> <i>libibcommon-devel-*.el5.ppc64.rpm</i> <i>libibcommon-static-*.el5.ppc64.rpm</i> <i>libibmad-*.el5.ppc64.rpm</i> <i>libibmad-devel-*.el5.ppc64.rpm</i> <i>libibmad-static-*.el5.ppc64.rpm</i> <i>libibumad-*.el5.ppc64.rpm</i> <i>libibumad-devel-*.el5.ppc64.rpm</i> <i>libibumad-static-*.el5.ppc64.rpm</i> <i>libibverbs-*.el5.ppc64.rpm</i> <i>libibverbs-devel-*.el5.ppc64.rpm</i> <i>libibverbs-static-*.el5.ppc64.rpm</i> <i>libibverbs-utils(it is used to ship ibv_* commands and depends on 32bit IB libraries) 64bit rpm is not available in RedHatEL5.3. Please install 32bit IB libraries also if user needs both ibv_* commands and the 64bit libraries.</i>
<b>libehca (for Galaxy1/ Galaxy2 support)</b>	<b>32bit</b>	<i>libehca-*.el5.ppc.rpm</i> <i>libehca-static-*.el5.ppc.rpm</i>
	<b>64bit</b>	<i>libehca-*.el5.ppc64.rpm</i> <i>libehca-static-*.el5.ppc64.rpm</i>
<b>libmthca (for Mellanox InfiniHost support)</b>	<b>32bit</b>	<i>libmthca-*.el5.ppc.rpm</i> <i>libmthca-static-*.el5.ppc.rpm</i>
	<b>64bit</b>	<i>libmthca-*.el5.ppc64.rpm</i> <i>libmthca-static-*.el5.ppc64.rpm</i>
<b>libmlx4 (for Mellanox ConnectX support)</b>	<b>32bit</b>	<i>libmlx4-*.el5.ppc.rpm</i> <i>libmlx4-static-*.el5.ppc.rpm</i>
	<b>64bit</b>	<i>libmlx4-*.el5.ppc64.rpm</i> <i>libmlx4-static-*.el5.ppc64.rpm</i>

On SLES:

Platforms	Driver/Library
-----------	----------------

<b>SLES11</b>	<i>ofed-1.4.0-3.4.ppc64.rpm</i>
<b>SLES10</b>	<i>libcxgb3-64bit-*.ppc.rpm</i> <i>libcxgb3-devel-*.ppc.rpm</i> <i>libcxgb3-devel-64bit-*.ppc.rpm</i> <i>libehca-*.ppc.rpm</i> <i>libehca-64bit-*.ppc.rpm</i> <i>libehca-devel-*.ppc.rpm</i> <i>libehca-devel-64bit-*.ppc.rpm</i> <i>libibcm-*.ppc.rpm</i> <i>libibcm-64bit-*.ppc.rpm</i> <i>libibcm-devel-*.ppc.rpm</i> <i>libibcm-devel-64bit-*.ppc.rpm</i> <i>libibcommon-*.ppc.rpm</i> <i>libibcommon-64bit-*.ppc.rpm</i> <i>libibcommon-devel-*.ppc.rpm</i> <i>libibcommon-devel-64bit-*.ppc.rpm</i> <i>libibmad-*.ppc.rpm</i> <i>libibmad-64bit-*.ppc.rpm</i> <i>libibmad-devel-*.ppc.rpm</i> <i>libibmad-devel-64bit-*.ppc.rpm</i> <i>libibumad-*.ppc.rpm</i> <i>libibumad-64bit-*.ppc.rpm</i> <i>libibumad-devel-*.ppc.rpm</i> <i>libibumad-devel-64bit-*.ppc.rpm</i> <i>libibverbs-*.ppc.rpm</i> <i>libibverbs-64bit-*.ppc.rpm</i> <i>libibverbs-devel-*.ppc.rpm</i> <i>libibverbs-devel-64bit-*.ppc.rpm</i> <i>libipathverbs-*.ppc.rpm</i> <i>libipathverbs-64bit-*.ppc.rpm</i> <i>libipathverbs-devel-*.ppc.rpm</i> <i>libipathverbs-devel-64bit-*.ppc.rpm</i> <i>libmlx4-*.ppc.rpm</i> <i>libmlx4-64bit-*.ppc.rpm</i> <i>libmlx4-devel-*.ppc.rpm</i> <i>libmlx4-devel-64bit-*.ppc.rpm</i>

*libmthca-\*.ppc.rpm*  
*libmthca-64bit-\*.ppc.rpm*  
*libmthca-devel-\*.ppc.rpm*  
*libmthca-devel-64bit-\*.ppc.rpm*  
*librdmacm-1.0.6-\*.ppc.rpm*  
*librdmacm-64bit-\*.ppc.rpm*  
*librdmacm-devel-\*.ppc.rpm*  
*librdmacm-devel-64bit-\*.ppc.rpm*  
*libsdp-\*.ppc.rpm*  
*libsdp-64bit-\*.ppc.rpm*  
*libsdp-devel-\*.ppc.rpm*  
*libsdp-devel-64bit-\*.ppc.rpm*  
*mpi-selector-\*.ppc.rpm*  
*mstflint-\*.ppc.rpm*  
*mvapich2-\*.ppc.rpm*  
*mvapich2-64bit-\*.ppc.rpm*  
*mvapich2-devel-\*.ppc.rpm*  
*mvapich2-devel-64bit-\*.ppc.rpm*  
*ofed-1.3-\*.ppc.rpm*  
*ofed-cxgb3-NIC-kmp-ppc64-\*.ppc.rpm*  
*ofed-doc-\*.ppc.rpm*  
*ofed-kmp-ppc64-\*.ppc.rpm*  
*open-iscsi-\*.ppc.rpm*  
*opensm-\*.ppc.rpm*  
*opensm-64bit-\*.ppc.rpm*  
*opensm-devel-\*.ppc.rpm*  
*opensm-devel-64bit-\*.ppc.rpm*  
*perftest-\*.ppc.rpm*  
*qlvnictools-\*.ppc.rpm*  
*rds-tools-\*.ppc.rpm*  
*release-notes-as-\*.ppc.rpm*  
*ruby-\*.ppc.rpm*  
*sdpnetstat-\*.ppc.rpm*  
*srptools-\*.ppc.rpm*  
*tvflash-\*.ppc.rpm*



## 8. Add this script to postscripts

```
chtab node=c890f11ec01 postscripts.postscripts=otherpkgs,configiba
```

*Note: postscript otherpkgs is used to install IB libraries/drivers.  
please include other postscripts what you need also.*

9. Now all the preparation work for IB configuration has been done, user can trigger updatenode command to update the nodes or continue to install the nodes.

Note: As a sample postscript, the netmask is set to default value: 255.255.0.0 and gateway is set to "X.X.255.254". If the IB interface name is not a simple combination of short hostname and ibX or netmask and gateway cannot meet user's requirement, user can modify this scripts by himself, like the example below:

The node short hostname is 890f11ec01-en, and the IB interface name is 890f11ec01-ib0, c890f11ec01-ib1, etc. The user needs to modify

```
my $hostname = "$ENV{NODE}-${nic}";  
to  
my $fullname = `echo $ENV{NODE} | cut -c 1-11`;  
chomp($fullname);  
my $hostname = "$fullname-${nic}";
```

It is assumed every node have two IB adapters, if only one adapter is available on each node, please modify script as following:

```
my @nums = (0..3);  
to  
my @nums = (0..1);
```

## 1.1. xDSH support IB switches

1. A new device configuration file on management node is introduced to provide special characteristics for xdsh command. The device configuration located in /var/opt/xcat/<DevicePath>/config.

The "<DevicePath>" is parsed from the attribute value of "--devicetype" flag (See description in item 4 below) or environment variable "DEVICETYPE" in xdsh commands invocation. If the devicetype for Qlogic switch is "IBSwitch::Qlogic" then the device configuration file must be found in the following directory:

```
/var/opt/xcat/IBSwitch/Qlogic/config
```

Below is an example of the device configuration file:

```
# Qlogic switch device configuration  
[main]  
ssh-setup-command=sshKey add  
[xdsh]  
pre-command=NULL  
post-command=showLastRetcode -brief
```

Below is the explanation:

- ssh-setup-comand

Specify the ssh key appending command supported by device specified. If this entry is not provided, xCAT uses default ways for HMC and IVM-managed devices to write ssh keys of Management Nodes.

- pre-command

Specify the pre-execution commands before remote command. For example, users might want to export some environment variables before executing real commands. If the value of this entry is assigned “NULL”, it means no pre-execution commands are needed.

For example, Qlogic Switch doesn't support environment variable, the 'pre-command' is assigned with “NULL” to disable environment variables usage.

If no entry is provided, the default behavior to export proper environment variables at the beginning of 'real' remote command will be used.

- post-command

Specify the built-in command provided by device specified to show the last command execution result. Qlogic Switch provides “showLastRetcode -brief” to display a numeric return code of last command execution.

If the value of this entry is assigned “NULL”, it means no post-command is used.

If no entry is provided, the default behavior to run “echo \$?” is used to dump return code of last command execution.

2.The default connection way on AIX is rsh, change rsh to ssh on AIX since ssh is only acceptable for automatically login setup.

```
chtab key=useSSHonAIX site.key=yes
```

3.Define IB switches as a node, this is required by xdsh which only support the input as a node.

```
mkdef -t node -o c890f11ec01 groups=all nodetype=switch
```

4.The xdsh command is used to concurrently run commands on multiple hardware devices. A new flag '--devicetype' and environment variable 'DEVICETYPE' is supported by xdsh. The '--devicetype' flag (env 'DEVICETYPE') is used to specify the type of target device, which is parsed to corresponding device configuration file location. xdsh will read this configuration file and use the specified commands for remote command execution. You must use the correct userid for your device.

xdsh required user to config ssh logging without password firstly. xdsh has provided the [-K ssh setup] option to do this automatically.

For example:

```
xdsh c890f11ec01 -K -w admin --devicetype IBSwitch::Qlogic
```

```
Enter the password for the userid on the node where the ssh keys will be updated.  
/usr/bin/ssh setup is complete.  
return code = 0
```

5. After setup of the ssh keys for the login, the admin can run the commands on IB switches from the management node using xdsh.

Below is an example of using xdsh.

```
/opt/xcat/bin/xdsh c890f11ec01 -l admin --devicetype IBSwitch::Qlogic ls  
Or export DEVICETYPE=IBSwitch::Qlogic /opt/xcat/bin/xdsh c890f11ec01 -l  
admin ls
```

## 1.1. Sample Scripts

### 1.1.1. Annotatelog

annotatelog is a sample script to parse the QLogic log entries in file `/var/log/xCAT/errorlog/[xCAT management nodes]` on xCAT Management Node by subnet manager, IB node, chassis, FRU(Field-Replaceable Unit) or a particular node. This script is supported by both AIX and Linux management nodes. From xCAT's point of view, the log to analyze must be xCAT consolidated log, which means this log file must come from xCAT syslog/errorlog monitoring mechanism, such as `/var/log/xCAT/errorlog/[xCAT Management Nodes]` file. Since the log format is various, xCAT does not support other log files.

The syntax of the annotatelog command will be:

```
annotatelog -f log_file [-s start_time] [-e end_time]  
             { [-i -g guid_file -l link_file] [-S] [-c] [-u] | [-A -g guid_file -l link_file] }  
             { [-n node_list -g guid_file] [-E] }  
             [-h]
```

#### **-f log\_file**

Specifies a log file fullpath name to analyze; must be xCAT consolidated log got from Qlogic HSM or ESM.

#### **-s start\_time**

Specifies the start time for analysis, where the **start\_time** variable has the format `ddmmyyhh:mm:ss` (day, month, year, hour, minute, and second), if it is not specified, annotatelog will parse the log file from the beginning.

#### **-e end\_time**

Specifies the end time for analysis, where the **end\_time** variable has the format ddmmyyhh:mm:ss (day, month, year, hour, minute, and second), if it is not specified, annotatelog will parse the log file to the end.

**-l link\_file**

Specifies a link file fullpath name, which concatenates all '/var/opt/iba/analysis/baseline/fabric\*links' files from all fabric management nodes.

**-g guid\_file**

Specifies a guid file fullpath name, which has a list of GUIDs as obtained from the "getGuids" script.

**-E**

Annotate with node ERRLOG\_ON and ERRLOG\_OFF information. This can help determine if a disappearance was caused by a node disappearing. It is for AIX nodes only and should be used with -n or -i flag

**-S**

Sort the log entries by subnet manager only.

**-i**

Sort the log entries by IB node only.

**-c**

Sort the log entries by chassis only.

**-u**

Sort the log entries by FRU only.

**-A**

Output the combination of -i, -S, -c and -u. It should be used with -g and -l flags.

**-n node\_list**

Specifies a comma-separated list of xCAT Managed Node host names, IP addresses to look up in log entries, it should be used with -g flag.

**-h**

Display usage information.

### 1.1.2. getGuids

getGuids is a sample script to get GUIDs for Infiniband Galaxy HCAs (Host Channel Adapter) and their ports from xCAT Management Nodes. It needs to be run from xCAT Management Node, and will use xdsh to all the xCAT Managed Nodes so that it could use ibstat command on AIX system or ibv\_devinfo command on Linux system to get the information about the IB devices.

The syntax of the getGuids command will be:

**getGuids [-h] [-f output\_file]**

**-f output\_file**

Specifies a file full path name that is used to save the GUIDs output.

**-h**

Display usage information.