

xCAT 2.0 Roadrunner Cookbook

04/08/2008

Table of Contents

| | | |
|----------------------------|---|----|
| 1.01.0 | Release Description..... | 3 |
| 2.02.0 | Installing the Management Node | 3 |
| 2.12.1 | Installing Fedora 8..... | 3 |
| 2.22.2 | Installing xCAT 2.0..... | 4 |
| 2.2.12.2.1 | Download xCAT2.0..... | 4 |
| 2.2.22.2.2 | Extract xCAT2.0..... | 4 |
| 2.2.32.2.3 | Get Fedora 8 OSS dependencies..... | 5 |
| 2.2.42.2.4 | Setup OSS repository..... | 6 |
| 2.2.52.2.5 | Install Dependencies..... | 6 |
| 2.2.62.2.6 | Install xCAT 2.0 software..... | 6 |
| 2.2.72.2.7 | Test xCAT installation..... | 6 |
| 2.2.82.2.8 | Update xCAT 2.0 software..... | 7 |
| 2.2.92.2.9 | Setup Yum for Fedora8 installs | 7 |
| 3.03.0 | xCAT Hierarchy using Service nodes..... | 8 |
| 3.13.1 | Switching to PostgreSQL Database..... | 8 |
| 3.23.2 | Define the service nodes in the database..... | 11 |
| 3.2.13.2.1 | Define Service Nodes and bmc in nodelist table..... | 11 |
| 3.2.23.2.2 | Define Service Nodes in noderes table..... | 11 |
| 3.2.33.2.3 | Define Service Nodes in ipmi table..... | 11 |
| 3.2.43.2.4 | Define Service Nodes and bmc in nodehm table..... | 12 |
| 3.2.53.2.5 | Define Service Nodes and bmc in nodetype table..... | 12 |
| 3.2.63.2.6 | Define Service Nodes in site table..... | 12 |
| 3.2.73.2.7 | Define Service Node OS and Profile attributes..... | 12 |
| 4.04.0 | Setup Services | 12 |
| 4.14.1 | Setup networks Table..... | 12 |
| 4.24.2 | Setup DNS..... | 12 |
| 4.34.3 | Setup AMM..... | 13 |
| 4.44.4 | Setup Conserver..... | 14 |
| 4.54.5 | Get MAC addresses..... | 14 |
| 4.64.6 | Setup DHCP..... | 14 |
| 4.74.7 | Startup TFTP..... | 15 |
| 5.05.0 | Define Compute Nodes in the Database..... | 15 |
| 5.15.1 | Setup the nodelist Table..... | 15 |
| 5.25.2 | Setup the noderes Table..... | 16 |
| 5.2.15.2.1 | Sample noderes table..... | 17 |
| 5.2.25.2.2 | Setting up which services run on the Service Nodes..... | 17 |

| | | | |
|------|---------|---|----|
| 5.3 | 5.3.3 | Setup nodetype table..... | 18 |
| | 5.3.1.1 | 5.3.1.1 Sample nodetype table..... | 18 |
| 5.4 | 5.4 | Setup passwords in passwd table..... | 18 |
| 5.5 | 5.5 | Setup deps Table for proper boot sequence..... | 18 |
| | 5.5.1.1 | 5.5.1.1 Sample deps table:..... | 18 |
| 6.0 | 6.0 | Build the service node stateless image..... | 19 |
| | 6.1 | 6.1 Install the Service Nodes..... | 20 |
| | 6.2 | 6.2 Test Service Node installation..... | 20 |
| 7.0 | 7.0 | iSCSI install QS22 blades..... | 20 |
| | 7.1 | 7.1 Build QS22 Stateless image..... | 21 |
| | 7.2 | 7.2 Install QS22 Stateless image | 22 |
| | 7.3 | 7.3 Update QS22 Stateless image..... | 23 |
| | 7.4 | 7.4 Build and Install QS22 Compressed Image..... | 24 |
| | 7.4.1 | 7.4.1 Build aufs..... | 24 |
| | 7.4.2 | 7.4.2 Generate the compressed image..... | 24 |
| | 7.4.3 | 7.4.3 Pack and install the compressed image..... | 25 |
| | 7.4.4 | 7.4.4 Check Memory Usage..... | 25 |
| | 7.4.5 | 7.4.5 Switch to iSCSI for more setup..... | 26 |
| 8.0 | 8.0 | Build LS21 Stateless image..... | 26 |
| | 8.1 | 8.1 Update LS21 Stateless image..... | 28 |
| | 8.2 | 8.2 Build and Install LS21 Compressed Image..... | 28 |
| | 8.2.1 | 8.2.1 Build aufs..... | 29 |
| | 8.2.2 | 8.2.2 Generate and pack the compressed image..... | 29 |
| | 8.2.3 | 8.2.3 Install the image..... | 30 |
| | 8.2.3.1 | 8.2.3.1 Check memory usage:..... | 30 |
| 9.0 | 9.0 | Service Node to Compute Node ssh setup..... | 30 |
| 10.0 | 10.0 | Building image for 64K pages..... | 31 |
| | 10.1 | 10.1 Rebuild aufs..... | 32 |
| | 10.2 | 10.2 Test unsquashed:..... | 33 |
| | 10.2.1 | 10.2.1 Check memory..... | 33 |
| | 10.3 | 10.3 Test squash..... | 34 |
| | 10.3.1 | 10.3.1 Check memory..... | 34 |
| | 10.4 | 10.4 Switch back to 4K pages..... | 34 |
| 11.0 | 11.0 | Installing OpenLDAP..... | 35 |
| | 11.1 | 11.1 Setup LDAP Server..... | 35 |
| | 11.1.1 | 11.1.1 Install the LDAP rpms | 36 |
| | 11.1.2 | 11.1.2 Configure LDAP..... | 37 |
| | 11.1.3 | 11.1.3 Migrate Users..... | 38 |
| | 11.2 | 11.2 Setup LDAP Client..... | 38 |
| | 11.2.1 | 11.2.1 Install LDAP into the image..... | 38 |
| | 11.2.2 | 11.2.2 Update the ldap configuration | 39 |
| | 11.2.3 | 11.2.3 Build the image and install | 40 |
| 12.0 | 12.0 | Setup Hierarchical LDAP..... | 40 |
| 13.0 | 13.0 | Install Torque..... | 40 |
| | 13.1 | 13.1 Setup Torque Server..... | 40 |
| | 13.2 | 13.2 Configure Torque..... | 41 |

| | | |
|--------------------------|--|----|
| 13.3 | 13.3 Define Nodes..... | 41 |
| 13.4 | 13.4 Setup and Start Service..... | 42 |
| 13.5 | 13.5 Install pbstop..... | 42 |
| 13.6 | 13.6 Install Perl Curses for PBS top..... | 42 |
| 13.7 | 13.7 Create a Torque default queue..... | 42 |
| 13.8 | 13.8 Setup Torque Client (x86_64 only)..... | 43 |
| 13.8.1 | 13.8.1 Install Torque..... | 43 |
| 13.8.2 | 13.8.2 Configure Torque..... | 43 |
| 13.8.2.1 | 13.8.2.1 Setup Access..... | 43 |
| 13.8.2.2 | 13.8.2.2 Setup node to node ssh for root | 44 |
| 13.8.3 | 13.8.3 Pack and Install image..... | 44 |
| 14.0 | 14.0 Setup Moab..... | 44 |
| 14.1 | 14.1 Install Moab..... | 45 |
| 14.2 | 14.2 Configure Moab..... | 45 |
| 14.2.1 | 14.2.1 Start Moab..... | 46 |
| 15.0 | 15.0 References..... | 46 |

1.0 Release Description

xCAT 2.0 is a complete rewrite of xCAT 1.2/1.3 implementing a new architecture. See xCAT 2.0 Beta Cookbook for more details about the 2.0 product:

<http://xcat.sourceforge.net/xCAT2.0.pdf> All commands are client/server, authenticated, logged and policy driven. The clients can be run on any OS with Perl, including Windows. The code has been completely rewritten in Perl, and table data is now stored in a relational database.

2.0 Installing the Management Node

2.1 Installing Fedora 8

Ensure that your networks are setup correctly.

1. Get Fedora ISOs and place in /root/xcat2:

```
mkdir /root/xcat2
cd /root/xcat2
wget
```

ftp://download.fedora.redhat.com/pub/fedora/linux/releases/8/Fedora/x86_64/iso/Fedora-8-x86_64-DVD.iso

```
wget ftp://download.fedora.redhat.com/pub/fedora/linux/releases/8/Fedora/ppc/iso/Fedora-8-ppc-DVD.iso
```

2. Create YUM repository:

```
mkdir /root/xcat2/fedora8  
mount -r -o loop /root/xcat2/Fedora-8-x86_64-DVD.iso /root/xcat2/fedora8
```

```
cd /etc/yum.repos.d  
mkdir ORIG  
mv fedora*.repo ORIG
```

Create fedora.repo with contents:

```
[fedora]  
name=Fedora $releasever - $basearch  
baseurl=file:///root/xcat2/fedora8  
enabled=1  
gpgcheck=0
```

3. Install createrepo
yum install createrepo

2.2 Installing xCAT 2.0

2.2.1 Download xCAT2.0

```
cd /root/xcat2  
wget http://xcat.sourceforge.net/yum/core-rpms-snap.tar.bz2  
wget http://xcat.sourceforge.net/yum/dep-snap/rh5/x86\_64
```

2.2.2 Extract xCAT2.0

```
cd /root/xcat2  
tar jxvf core-rpms-snap.tar.bz2  
tar jxvf dep-rpms-snap.tar.bz2
```

2.2.3 Get Fedora 8 OSS dependencies

```
cd /root/xcat2/dep-snap
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/perl-Net-SNMP-5.2.0-1.fc8.1.noarch.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/perl-XML-Simple-2.17-1.fc8.noarch.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/perl-Crypt-DES-2.05-4.fc7.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/net-snmp-perl-5.4.1-4.fc8.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/ksh-20070628-1.1.fc8.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/perl-IO-Socket-INET6-2.51-2.fc8.1.noarch.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/dhcp-3.0.6-10.fc8.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/syslinux-3.36-7.fc8.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/mtools-3.9.11-2.fc8.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/expect-5.43.0-9.fc8.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/perl-DBD-SQLite-1.12-2.fc8.1.x86\_64.rpm
```

```
wget
```

```
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86\_64/os/Packages/perl-Expect-1.20-1.fc8.1.noarch.rpm
```

wget

http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Everything/x86_64/os/Packages/perl-IO-Tty-1.07-2.fc8.1.x86_64.rpm

wget

http://mirrors.usc.edu/pub/linux/distributions/fedora/linux/releases/8/Everything/x86_64/os/Packages/scsi-target-utils-0.0-1.20070803snap.fc8.x86_64.rpm

2.2.4 Setup OSS repository

```
cd /root/xcat2/dep-snap
```

```
./mklocalrepo.sh
```

```
createrepo $PWD
```

```
yum clean metadata
```

2.2.5 Install Dependencies

```
yum install rpm-build perl-IO-Socket-SSL perl-Net-SSLeay \
```

```
perl-Digest-HMAC perl-Digest-SHA1 bind dhcp \
```

```
perl-Expect perl-IO-Socket-INET6 syslinux \
```

```
perl-Net-SNMP perl-XML-Simple \
```

```
net-snmp-perl expect ksh atftp conserver \
```

```
fping ipmitool perl-DBD-SQLite
```

2.2.6 Install xCAT 2.0 software

```
cd /root/xcat2/core-snap
```

```
./mklocalrepo.sh
```

```
createrepo $PWD
```

```
yum clean metadata
```

```
yum install xCAT-2.0
```

2.2.7 Test xCAT installation

```
source /etc/profile.d/xcat.sh
```

```
tabdump site
```

2.2.8 Update xCAT 2.0 software

If you need to update the xCAT2.0 rpms later with a new copy:

```
cp <new rpms> /root/xcat2/core-snap
cd /root/xcat2/core-snap
./mklocalrepo.sh
createrepo $PWD
yum update
```

If you have a service node stateless image, don't forget to update the image with the new xCAT rpms. See xxxxxx building stateless images for service nodes.

```
cp -pf /etc/yum.repos.d/*.repo
/install/netboot/fedora9/x86_64/service/rootimg/etc/yum.repos.d
yum --installroot=/install/netboot/fedora8/x86_64/service/rootimg update
```

2.2.9 Setup Yum for Fedora8 installs

```
umount /root/xcat2/fedora8
chtab key=installdir site.value=/install
cd /root/xcat2
copycds Fedora-8-x86_64-DVD.iso
copycds Fedora-8-ppc-DVD.iso
Edit /etc/yum.repos.d/fedora.repo
```

```
Change
baseurl=file:///tmp/fedora8
to
baseurl=file:///install/fedora8/x86_64
```

3.0 xCAT Hierarchy using Service nodes

In large clusters it is desirable to have more than one node (the Management Node) handle the installation of the compute nodes. We call these additional nodes service nodes. You can have one or more service nodes setup to install groups of compute nodes.

The service nodes need to communicate with the xCAT2.0 database on the Management Node and run xCAT command to install the nodes. The service node will be installed with the xCAT code and required the PostgreSQL Database be setup instead of SQLite Default database. PostgreSQL allows a client to be setup on the service node such that the service node can access (read/write) the database on the Management Node (Master Node) from the service node.

If you do not plan on using service nodes, you can skip this section 3 and continue to use the SQLite Default database setup during the installation.

3.1 Switching to PostgreSQL Database

To setup the postgresql database on the Management Node follow these steps.

This example assumes:

192.168.0.1: ip of master

xcatdb: database name

xcatadmin: database role (aka user)

cluster: database password

192.168.0.10 & 192.168.0.11 (service nodes)

Substitute your address and desired userid , password and database name as appropriate.

The following rpms should be installed from the Fedora8 media on the Management Node (and service node when installed). These are required for postgresql.

1. yum install perl-DBD-Pg postgresql-server postgresql
2. Initialize the database :
service postgresql initdb
3. service postgresql start
4. su – postgres
5. -bash-3.1\$ createuser -P xcatadmin
Enter password for new role: cluster

Enter it again: cluster
Shall the new role be a superuser? (y/n) n
Shall the new role be allowed to create databases? (y/n) n
Shall the new role be allowed to create more new roles? (y/n) n

6. \$ createdb -O xcatadmin xcatdb

7. \$ exit

8. cd /var/lib/pgsql/data/

9. Edit the hba configuration file:

vi pg_hba.conf

#lines should look like:

local all all ident sameuser

IPv4 local connections:

host all all 127.0.0.1/32 md5

host all all 192.168.0.1/32 md5

host all all 192.168.0.10/32 md5

host all all 192.168.0.11/32 md5 where 192.168.0.10 and 11 are service nodes.

10. vi postgresql.conf

11. set listen_addresses to '*':

listen_addresses = '*' This allows remote access. **Note:Be sure and uncomment the line**

12. service postgresql restart

13. Backup your data to migrate to the new database

#mkdir -p ~/xcat-dbback

dumpxCATdb -p ~/xcat-dbback

14. /etc/sysconfig/xcat should contain these lines, substitute your cluster facing address for 192.168.0.1, and user and password are xcatadmin cluster in this instance

XCATCFG='Pg:dbname=xcatdb;host=192.168.0.1|xcatadmin|cluster'

export XCATCFG

XCATROOT=/opt/xcat

export XCATROOT

15. copy /etc/sysconfig/xcat to /install/postscripts/sysconfig/xcat for installation on the service nodes.

16. /etc/xcat/cfgloc should contain the following line, again substituting your info. This points the xCAT database access code to the new database.

Pg:dbname=xcatdb;host=192.168.0.1|xcatadmin|cluster

17. copy `/etc/xcat/cfgloc` to `/install/postscripts/etc/xcat/cfgloc` for installation on the service nodes.
18. `chmod 700 /etc/sysconfig/xcat` and `/etc/xcat/cfgloc`
19. `./etc/sysconfig/xcat` #read the text into the current shell
20. You can add `./etc/sysconfig/xcat` to a setup shell script in `/etc/profile.d`, so the `XCATROOT` and `XCATCFG` environment variables are setup when you login.

21. Start the `xcatd` daemon using the postgresql database
`service xcatd restart`
22. Restore your database: `restorexCATdb -p ~/xcat-dbback` to the postgresql database

23. Need to update the policy table:
Run this command to get correct Master node name known by ssl:
`openssl x509 -text -in /etc/xcat/cert/server-cert.pem -noout|grep Subject`
Subject: CN=mgt.cluster
Subject Public Key Info:
X509v3 Subject Key Identifier:

24. Update the policy table with `mgt.cluster` output from the command:
`chtab priority=5 policy.name=<mgt.cluster> policy.rule=allow`. Note this name must be an MN name that is known by the service nodes.

25. Check the database for the following settings:

```
[root@mn20 ~]# tabdump site
```

```
#key,value,comments,disable
```

```
"xcatiport","3002",,
```

```
"nameservers","11.16.0.1",,
```

```
"forwarders","9.114.8.1,9.114.8.2",,
```

```
"xcatdport","3001",,
```

```
"domain","foobar.com",,
```

```
"master","11.16.0.1",, where the Master node is the name or ip address known by the service nodes.
```

```
[root@mn20 ~]# tabdump policy
```

```
#priority,name,host,commands,noderange,parameters,time,rule,comments,disable
```

```
"1","root" ,,,,,,"allow" ,,
"5","mn20" ,,,,,,"allow" ,, where mn20 is the output of step 26.
"2" ,,, "getbmconfig" ,,,,"allow" ,,
"3" ,,, "nextdestiny" ,,,,"allow" ,,
"4" ,,, "getdestiny" ,,,,"allow" ,,
```

```
26. chkconfig postgresql on
27. service postgresql restart
```

3.2 Define the service nodes in the database

For this example, we have two service nodes rra000 and rrb000. To add the service nodes to the database run the following commands to add and update the service nodes' attributes in the site, nodelist and noderes tables. Note: service nodes are required to be defined with group “service”. The commands below are using the group “service” to update all service nodes.

Note: For table attribute definitions run `tabdump -d <table name>`

3.2.1 Define Service Nodes and bmc in nodelist table

```
nodeadd rra000,rrb000 groups=service,ipmi, all
nodeadd rra000bmc,rrb000bmc groups=bmc,ipmi,all
```

3.2.2 Define Service Nodes in noderes table

```
chtab node=service noderes.netboot=pxe
chtab node=service noderes.servicenode="11.16.0.1"
chtab node=service noderes.tftpserver="11.16.0.1"
chtab node=service noderes.xcatmaster="11.16.0.1"
chtab node=service noderes.serialport=1
chtab node=service noderes.service="11.16.0.1"
```

3.2.3 Define Service Nodes in ipmi table

```
nodech rra000 ipmi.bmc=rra000bmc ipmi.userid=USERID
ipmi.password=PASSWORD
nodech rrb000 ipmi.bmc=rrb000bmc ipmi.userid=USERID
ipmi.password=PASSWORD
```

3.2.4 Define Service Nodes and bmc in nodehm table

```
chtab node=service nodehm.cons=ipmi
```

```
chtab node=service nodehm.mgt=ipmi nodehm.serialspeed=19200  
nodehm.serialflow=hard
```

```
chtab node=bmc nodehm.mgt=ipmi
```

3.2.5 Define Service Nodes and bmc in nodetype table

```
chtab node=service nodetype.arch=x86_64 nodetype.os=fedora8  
nodetype.nodetype=osi
```

```
chtab node=bmc nodetype.nodetype=rsa
```

3.2.6 Define Service Nodes in site table

```
chtab key=defserialport site.value=1
```

```
chtab key=defserialspeed site.value=19200
```

```
chtab key=xcatservers site.value=rra000,rrb000
```

3.2.7 Define Service Node OS and Profile attributes

```
chtab node=service nodetype.os=fedora8
```

```
chtab node=service noderes.primarynic=eth0 noderes.installnic=eth0
```

```
chtab node=service nodetype.profile=service
```

4.0 Setup Services

4.1 Setup networks Table

All networks in the cluster must be defined in the networks table.

4.2 Setup DNS

Set nameserver, forwarders and domain in the site table

```
chtab key=nameservers site.value=192.168.100.1 (IP of mgmt node)
```

```
chtab key=forwarders site.value=172.16.0.1 (how to get to other DNS)
```

```
chtab key=domain site.value=foobar.com
```

Edit /etc/hosts:

```
127.0.0.1    localhost.localdomain localhost
::1         localhost6.localdomain6 localhost6
192.168.2.100 b7-eth0
192.168.100.1 b7
192.168.100.10 blade1
192.168.100.11 blade2
192.168.100.12 blade3
172.30.101.133 amm3
```

Run:

```
makenetworks
```

```
makedns
```

```
setup /etc/resolv.conf:
search foobar.com
nameserver 192.168.100.1
```

Start dns:

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

4.3 Setup AMM

Note: xCAT will be providing a script to replace this manual process.

```
telnet amm3
env -T mm[1]
users -1 -ap sha -ppw PASSWORD
users -1 -at set
```

As one line copy your id_rsa.pub key from the \$ROOTHOME/.ssh/id_rsa.pub file.

```
users -1 -pk -add ssh-rsa
AAAAB3NzaC1yc2EAAAABIwAAAQEA0u4zf9ULqp5jsZPiVlmcg8TWbPrrIyOK
+bMbHPmId0OEQvs6Opc12XqC4VF6POH8zEu6/YmpPphuDqhOmjkou/TXxHgZJ
```

```
KQmZ/gFK7Fr9dFzbwA37eE0edeOK4WolwNZgH7t
+4Bm1fJ1sjELVIR1CjFSm59c6Fts83NKIeU6wuhEOzYG1UywyW1Aj/0rSLOk1pS
Fklhu9yXwt9RNVyQva7KKFhXFS51WaFRjyjEMU1Mc/AKaHYnNdehVSm3Bpks
dMIkOVC36/VCXdwqEZWkV0m1pgCIM4K8CPfQUyuP3iaBep2hLA6o8f4bwrXM
XAckrORWCKzFuiV3QoBCAJKxKPQ== root@mgt.cluster
```

NOTE: If you get error with -add type reset and try again. MM Bugs.

Test with:

```
ssh USERID@amm3 exit
```

TIP to update firmware:

```
Put CNETCMUS.pkt in /tftpboot
```

```
telnet AMM
```

```
env -T mm[1]
```

```
update -v -i TFTP_SERVER_IP -l CNETCMUS.pkt
```

TIP for SOL to work best telnet to nortel switch and type:

```
/cfg/port int1/gig/auto off, for each port.
```

4.4 Setup Conserver

```
makeconservercf
```

```
service consver stop
```

```
service consver start
```

Test a few nodes with rpower and wcons

4.5 Get MAC addresses

```
rinv all macs |
```

```
perl -pi -e 's/([^\:]*):.*?ress (\d): (00(:[0-9A-F]{2}){5})/nodech \1 mac.mac=\3 #\2/' |
grep \#1
```

tabdump mac to verify mac addresses in table.

4.6 Setup DHCP

Setup dynamic range for your networks, for example:

```
chtab net=192.168.100.0 networks.dynamicrange=192.168.100.200-192.168.100.250
```

Define dhcp interfaces in site table:

```
chtab key=dhcpinterfaces site.value=eth1
```

Start dhcp:

```
service dhcpd restart
```

Create dhcp leases files

```
makedhcp -n
```

```
service dhcpd restart
```

4.7 Startup TFTP

```
mknb x86_64
```

```
service tftpd restart
```

5.0 Define Compute Nodes in the Database

5.1 Setup the nodelist Table

The nodelist table contains a node definition for each node in the cluster. We have provided a script to automate these definitions for the RR cluster.

`/opt/xcat/share/xcat/tools/mkrrnodes` will allow you to automatically define as many nodes as you would like to and setup nodegroups needed to manage those nodes. See `man mkrrnodes`.

For example :

Running `mkrrnodes` will define the following nodes with the assigned groups in the nodelist table. These nodegroups will be used in additional xCAT Table setup so that an entry does not have to be made for every node. You can add any additional nodegroups that you would like to define with the `tabedit` command.

```
/opt/xcat/share/xcat/tools/mkrrnode -C b -R 047,048
```

adds to the nodelist table the following entries:

```
"rrb047a", "rrb047,ls21,cub,opteron,opteron-cub,compute,tb,all,rack06" ,,,  
"rrb047b", "rrb047,qs22,cub,cell,cell-cub-b,compute,tb,all,rack06" ,,,  
"rrb047c", "rrb047,qs22,cub,cell,cell-cub-c,compute,tb,all,rack06" ,,,  
"rrb048a", "rrb048,ls21,cub,opteron,opteron-cub,compute,all,tb,rack06" ,,,  
"rrb048b", "rrb048,qs22,cub,cell,cell-cub-b,compute,tb,all,rack06" ,,,  
"rrb048c", "rrb048,qs22,cub,cell,cell-cub-c,compute,tb,all,rack06" ,,,
```

5.2 Setup the noderes Table

The noderes table will define for the node or nodegroup, the service node used to service the node or group, the type of network booting supported, the node which is the tftpserver, dhcpserver,etc as known by the node.

If you are using service nodes, for each node or nodegroup defined in the noderes table change the service node attribute in the noderes table to point to the name or ip address of it's service node.

So for nodes in group rrb048, assign rrb000 service node to the node group and the xcatmaster will be the address that the node knows the service node by.

```
chtab node=opteron-cub noderes.servicenode=rrb000 noderes.xcatmaster=rrb000
```

```
chtab node=cell-cub-b noderes.servicenode=rrb000 noderes.xcatmaster=rrb000
```

```
chtab node=cell-cub-c noderes.servicenode=rrb000 noderes.xcatmaster=rrb000
```

Note: we are using 3 different nodegroups here because there will be different entries in the noderes table for the cell blades vs the opteron.

Define the services to run on the servicenode for the node group, for example to setup tftpserver and nfsserver

```
chtab node=opteron-cub noderes.tftpserver=rrb000noderes.nfsserver=rrb000
```

```
chtab node=cell-cub-b noderes.tftpserver=rrb000 noderes.nfsserver=rrb000
```

```
chtab node=cell-cub-c noderes.tftpserver=rrb000 noderes.nfsserver=rrb000
```

Whether or not you are using Service Nodes:

Define the type of network booting supported by this type of node (pxe,yaboot). If no service node, the xcatmaster is the Master Node.

```
chtab node=opteron-cub noderes.netboot=pxe noderes.master="11.16.0.1"
```

```
chtab node=cell-cub-b noderes.netboot=yaboot noderes.master="11.16.0.1"
```

```
chtab node=cell-cub-c noderes.netboot=yaboot noderes.master="11.16.0.1"
```


Define the network adapters that will be used for deployment.

```
chtab node=opteron-cub noderes.primarynic=eth0 noderes.installnic=eth0
```

```
chtab node=cell-cub-b noderes.primarynic=eth0 noderes.installnic=eth0
```

```
chtab node=cell-cub-c noderes.primarynic=eth0 noderes.installnic=eth0
```

5.2.1 Sample noderes table

Your noderes table will end up looking like this (if you use service nodes):

```
node,servicenode,netboot,tftpserver,nfsserver,monserver,kernel,initrd,kcmdline,nfsdir,serialport,installnic,primarynic,xcatmaster,current_osimage,next_osimage,comments,disable
"opteron-cub,
"11.17.0.1","pxe","11.17.0.1","11.17.0.1",,,,,,"1","eth0","eth0","11.18.0.1",,,,
"cell-cub-b",
"11.18.0.1","yaboot","11.18.0.1","11.18.0.1",,,,,,"0","eth0","eth0","11.18.0.1",,,,
"cell-cub-c",
"11.18.0.1","yaboot","11.18.0.1","11.18.0.1",,,,,,"0","eth0","eth0","11.18.0.1",,,,
```

5.2.2 Setting up which services run on the Service Nodes

Note: if in the noderes table you have an assigned servicenode for a node, and the field for the service (e.g nfsserver) is left blank, it is assumed that you want that service running on the defined service node. So you can either explicitly assign a service node to a node for any given service, or you can leave the fields blank and the service node assigned to the node will run all services for that node.

The settings for the services in the database will determine which services are setup on the service node. These services are setup when the xcatd daemon is started on the service node.

The services that are setup by xCAT on the service node are as follows:

- nfs (always setup)
- dns
- conserver
- tftp
- http (automatically installed)
- dhcp
- syslog (always setup)

5.3 Setup nodetype table

Define the OS and profile type for building the stateless image.

```
chtab node=opteron nodetype.os=fedora8 nodetype.profile=compute
```

```
chtab node=cell nodetype.os=fedora8 nodetype.profile=compute
```

```
chtab node=service nodetype.os=fedora8 nodetype.profile=service
```

5.3.1.1 Sample nodetype table

Your nodetype table will look something like this:

```
#node,os,arch,profile,nodetype,comments,disable
```

```
"service","fedora8","x86_64","service",,,
```

```
"opteron","fedora8","x86_64","compute",,,
```

```
"cell","fedora8","ppc64","compute",,,
```

5.4 Setup passwords in passwd table

Add needed passwords to the passwd table to support installs.

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

```
chtab key=blade passwd.password=PASSWORD passwd.username=USERID
```

```
chtab key=ipmi passwd.password=PASSWORD passwd.username=USERID
```

Note: PASSWORD (zero not letter O)

5.5 Setup deps Table for proper boot sequence

The following is an example of how you can setup the deps table to ensure the triblades boot up in the proper sequence:

Use tabedit deps to input the values.

Use rpower -d for triboot

5.5.1.1 Sample deps table:

```
#node,nodedep,msdelay,cmd,comments,disable
"opteron","|rr(.d+)a|rr($1)b,rr($1)c|","5000","on",,
"cell","|rr(.d+).|rr($1)a|","5000","off",,
```

6.0 Build the service node stateless image

The service node stateless images must contain not only the OS, but also xCAT2.0. In addition a number of files are added to the image to support the postgresql database access from the service node to the Management node, and ssh access to the nodes that the service nodes services. Note: the following example assumes you are building the stateless image on the Management Node.

1. Check the service node packaging to see if it has all the rpms required.

```
cd /opt/xcat/share/xcat/netboot/fedora/
```

```
vi service.exlist and service.pklist
```

To add packages:

```
echo vi >>service.pkglis
```

```
    echo dhcp >>service.pkglis
```

```
    echo atftp >>service.pkglis
```

```
    echo bind >>service.pkglis
```

```
    echo nfs-utils >>service.pkglis
```

Include things you may need, for example by editing service.exlist and remove the following line:

```
    ./usr/lib/perl5*
```

Edit compute.exlist, if necessary, adding lines to remove unnecessary rpms.

2. Run image generation:

```
cd /opt/xcat/share/xcat/netboot/fedora/
```

```
./genimage -i eth0 -n tg3,bnx2 -o fedora8 -p service
```

3. Install xCAT code into the service node image:

```
rm -f /install/netboot/fedora8/x86_64/service/rootimg/etc/yum.repos.d/*
```

```
cp /etc/yum.repos.d/fedora.repo
```

```
/install/netboot/fedora8/x86_64/service/rootimg/etc/yum.repos.d
```

```
yum --installroot=/install/netboot/fedora8/x86_64/service/rootimg install xCATsn
```

4. Update the service node image with the additional files needed for setting up keys and postgresql db when installed:

```
updateSNimage -p /install/netboot/fedora8/x86_64/service/rootimg
```

5. Add automatic configuration of eth1 adapter on the service node

```
chroot /install/netboot/fedora8/x86_64/service/rootimg
```

```
bash-3.2# chkconfig --add xcatd-hack
```

6. Pack the image

```
packimage -o fedora8 -p service -a x86_64
```

6.1 Install the Service Nodes

```
nodeset service netboot
```

```
rpower service boot
```

6.2 Test Service Node installation

ssh to the service node.

Check to see that the xcat daemon xcatd is running.

Run some database command on the service node, e.g tabdump site, nodels and see that the database can be accessed from the service node.

Check that /install and /tftpboot are mounted on the service node from the Management Node.

7.0 iSCSI install QS22 blades

```
yum install yaboot-xcat scsi-target-utils
```

```
chtab key=iscsidir site.value=/install/iscsi
```

Pick one of the QS22 blades for the iSCSI install

Note: make sure the root userid and password are in the iscsi table

```
chtab node=rrb047b iscsi.userid=root iscsi.password=cluster iscsi.server="11.16.0.1"
```

```
chtab node=rrb047b noderes.nfsserver="11.16.0.1" (MasterNode)
```

```
chtab node=rrb047b nodetype.os=fedora8 nodetype.profile=iscsi groups,=iscsi  
iscsi.server="11.16.0.1"
```

```
service tgtd restart
```

```
nodech rrb047b iscsi.file=
```

```
setupiscsidev -s8192 rrb047b
```

```
nodeset rrb047b install
```

NOTE: for reinstall:

```
chtab node=rrb047b nodetype.profile=iscsi
```

7.1 Build QS22 Stateless image

1. Logon to the node

```
ssh rrb047b
```

```
mkdir /install
```

```
mount 11.16.0.1:/install /install
```

2. Create fedora.repo:

```
cd /etc/yum.repos.d
```

```
rm -f *.repo
```

Put the following lines in /etc/yum.repos.d/fedora.repo:

```
[fedora]
```

```
name=Fedora $releasever - $basearch
```

```
baseurl=file:///install/fedora8/ppc64
```

```
enabled=1
```

```
gpgcheck=0
```

3. Test with: yum search gcc

4. Copy the executables and files needed from the Management Node:
cd /root
scp 11.16.0.1:/opt/xcat/share/xcat/netboot/fedora/genimage .
scp 11.16.0.1:/opt/xcat/share/xcat/netboot/fedora/geninitrd .
scp 11.16.01.:/opt/xcat/share/xcat/netboot/fedora/compute.ppc64.pkglist .
5. Generate the image:
./genimage -i eth0 -n tg3 -o fedora8 -p compute

NOTE: iSCSI, QS22, tg3, all slow, take a nap

7.2 Install QS22 Stateless image

On the Management Node:

1. Adding Service Node ssh keys
See 309 for how to add keys to the install image to be able to ssh from the Service Node to the compute nodes.
2. Edit fstab in the image

```
cd /install/netboot/fedora8/ppc64/compute/rootimg/etc  
cp fstab fstab.ORIG
```

Edit fstab:

Change:

```
devpts /dev/pts devpts gid=5,mode=620 0 0  
tmpfs /dev/shm tmpfs defaults 0 0  
proc /proc proc defaults 0 0  
sysfs /sys sysfs defaults 0 0
```

to:

```
proc      /proc          proc      rw 0 0
sysfs     /sys           sysfs     rw 0 0
devpts    /dev/pts       devpts    rw,gid=5,mode=620 0 0
#tmpfs    /dev/shm       tmpfs     rw 0 0
compute_ppc64  /           tmpfs     rw 0 1
none      /tmp           tmpfs     defaults,size=10m 0 2
none      /var/tmp       tmpfs     defaults,size=10m 0 2
```

3. Pack the image

```
packimage -o fedora8 -p compute -a ppc64
```

4. Install the image on all the QS22 blades

```
chtab node=cell nodetype.profile=compute nodetype.os=fedora8
nodeset cell netboot
rpower cell boot
```

7.3 Update QS22 Stateless image

NOTE: before YUM/RPM commands type:

```
rm /install/netboot/fedora8/ppc64/compute/rootimg/var/lib/rpm/__.db.00*
```

1. Update image using YUM

```
rm -f /install/netboot/fedora8/ppc64/compute/rootimg/etc/yum.repos.d/*
cp /etc/yum.repos.d/fedora.repo
/install/netboot/fedora8/ppc64/compute/rootimg/etc/yum.repos.d
```

Now install vi into the image:

```
yum --installroot=/install/netboot/fedora8/ppc64/compute/rootimg install vi
```

2. Update image using RPM

```
rpm --root /install/netboot/fedora8/ppc64/compute/rootimg -Uvh
/install/fedora8/ppc64/Packages/vim-minimal-7.1.135-1.fc8.ppc.rpm
```

3. Update the image by running genimage

```
Add packages to compute.ppc64.pkglint and rerun genimage
```

4. packimage -o fedora8 -p compute -a ppc64

7.4 Build and Install QS22 Compressed Image

On the QS22 blade:

```
yum install kernel-devel gcc squashfs-tools
```

On net connected node:

```
svn co http://xcat.svn.sf.net/svnroot/xcat/xcat-dep/trunk/aufs
```

7.4.1 Build aufs

```
cd aufs
```

```
tar jxvf aufs-2-6-2008.tar.bz2
```

```
cd aufs
```

```
mv include/linux/aufs_type.h fs/aufs/
```

```
cd fs/aufs/
```

```
patch -p1 < ../.././aufs-standalone.patch
```

```
chmod +x build.sh
```

```
./build.sh
```

```
# ls -lh aufs.ko
```

```
-rw-r--r-- 1 root root 3.5M 2008-03-10 14:20 aufs.ko
```

```
strip -g aufs.ko
```

```
cp aufs.ko /root
```

7.4.2 Generate the compressed image


```
cd /opt/xcat/share/xcat/netboot/fedora
```

```
./geninitrd -i eth0 -n tg3,squashfs,aufs,loop -o fedora8 -p compute -l $(expr 100 \* 1024 \* 1024)
```

7.4.3 Pack and install the compressed image

On the Management Node:

```
yum install squashfs-tools
```

```
packimage -a ppc64 -o fedora8 -p compute -m squashfs
```

```
chtab node=cell nodetype.profile=compute nodetype.os=fedora8
```

```
nodeset cell netboot
```

```
rpower cell boot
```

7.4.4 Check Memory Usage

```
# ssh left "echo 3 > /proc/sys/vm/drop_caches;free -m;df -h"
```

```
      total    used    free   shared  buffers   cached
Mem:    3961    99   3861      0      0      61
-/+ buffers/cache:    38   3922
Swap:    0      0      0

Filesystem      Size  Used Avail Use% Mounted on
compute_ppc64  100M 220K 100M  1% /
none            10M   0 10M  0% /tmp
none            10M   0 10M  0% /var/tmp
```

Max for / is 100M, but only 220K being used (down from 225M), but wheres the OS?

Look at cached. 61M compress OS image. 3.5x smaller

As files change in hidden OS then get copied to tmpfs (compute_ppc64) with a copy on write. To reclaim space reboot. The /tmp and /var/tmp is for MPI and other Torque and user related stuff. if 10M is too small you can fix it. To reclaim this space put in epilogue umount /tmp /var/tmp; mount -a..

Reboot cell as stateless, from mgmt to reclaim space:

Set profile to compute:

```
chtab cell nodetype.profile=compute nodetype.os=fedora8
```

```
nodeset cell netboot
```

```
xdsh cell reboot #be nice, iSCSI is still stateful, be kind to the state
```

7.4.5 Switch to iSCSI for more setup

To reboot rra047b as iscsi for more stateless setup fun:

```
nodech rra047b nodetype.profile=iscsi
```

```
nodeset rra047b iscsiboot
```

```
rpower rra047b boot
```

8.0 Build LS21 Stateless image

The LS21 image can be built on the Management Node since it is of the same architecture.

Note: For hierarchical support -- use noderes.servicenode for http image server

On the Management Node:

1. Check the compute node packaging to see if it has all the rpms required.

```
cd /opt/xcat/share/xcat/netboot/fedora/
```

```
vi compute.exlist and compute.pklist
```

To add packages:

```
echo vi >>compute.pklist
```

Include things you may need, for example by editing compute.exlist and remove the following line:

```
./usr/lib/perl5*
```

Edit compute.exlist, if necessary, adding lines to remove unnecessary rpms.

2. Run image generation:
`cd /opt/xcat/share/xcat/netboot/fedora/`
`./genimage -i eth0 -n tg3,bnx2 -o fedora8 -p compute`
3. Adding Service Node ssh keys
 See 309 for how to add keys to be able to ssh from the Service Node to the compute nodes.
4. Edit fstab in the image

```
cd /install/netboot/fedora8/x86_64/compute/rootimg/etc
cp fstab fstab.ORIG
```

Edit fstab:

Change:

```
devpts /dev/pts devpts gid=5,mode=620 0 0
tmpfs /dev/shm tmpfs defaults 0 0
proc /proc proc defaults 0 0
sysfs /sys sysfs defaults 0 0
```

to:

```
proc /proc proc rw 0 0
sysfs /sys sysfs rw 0 0
devpts /dev/pts devpts rw,gid=5,mode=620 0 0
#tmpfs /dev/shm tmpfs rw 0 0
compute_x86_64 / tmpfs rw 0 1
none /tmp tmpfs defaults,size=10m 0 2
none /var/tmp tmpfs defaults,size=10m 0 2
```

5. Package the image

```
packimage -o fedora8 -p compute -a x86_64
```

6. Install the image on all the LS21 blades

```
chtab node=opteron nodetype.profile=compute nodetype.os=fedora8
```

```
nodeset opteron netboot
```

```
rpower opteron boot
```

8.1 Update LS21 Stateless image

1. Update image using YUM

NOTE: before YUM/RPM commands type:

```
rm /install/netboot/fedora8/x86_64/compute/rootimg/var/lib/rpm/__db.00*
```

```
rm -f /install/netboot/fedora8/x86_64/compute/rootimg/etc/yum.repos.d/*
```

```
cp /etc/yum.repos.d/fedora.repo /install/netboot/fedora8/x86_64/compute/rootimg/etc/  
yum.repos.d
```

Now install vi into the image:

```
yum --installroot=/install/netboot/fedora8/x86_64/compute/rootimg install vi
```

2. Update image using RPM

```
rpm --root /install/netboot/fedora8/x86_64/compute/rootimg -Uvh blah.rpm
```

3. Repackage

```
packimage -o fedora8 -p compute -a x86_64
```

4. Install on all LS21 blades

```
nodeset opteron netboot
```

```
rpower opteron boot
```

8.2 Build and Install LS21 Compressed Image

On Management Node:

```
yum install kernel-devel gcc squashfs-tools
```

8.2.1 Build aufs

```
svn co http://xcat.svn.sf.net/svnroot/xcat/xcat-dep/trunk/aufs
```

```
cd aufs
```

```
tar jxvf aufs-2-6-2008.tar.bz2
```

```
cd aufs
```

```
mv include/linux/aufs_type.h fs/aufs/
```

```
cd fs/aufs/
```

```
patch -p1 < ../../aufs-standalone.patch
```

```
chmod +x build.sh
```

```
./build.sh
```

```
ls -lh aufs.ko
```

```
-rw-r--r-- 1 root root 3.2M 2008-02-27 13:09 aufs.ko
```

```
strip -g aufs.ko
```

```
cp aufs.ko /opt/xcat/share/xcat/netboot/fedora/
```

8.2.2 Generate and pack the compressed image

```
cd /opt/xcat/share/xcat/netboot/fedora
```

```
./geninitrd -i eth0 -n tg3,bnx2,squashfs,aufs,loop -o fedora8 -p service -l $(expr 100 \  
* 1024 \* 1024)
```

Pack the compressed image

```
packimage -a x86_64 -o fedora8 -p compute -m squashfs
```

NOTE: To unsquash:

```
cd /install/netboot/fedora8/x86_64/service
```

```
rm -f rootimg.sfs
packimage -a x86_64 -o fedora8 -p service -m cpio
```

NOTE: The -l and -t is the size of the / and /tmp,/var/tmp file systems in RAM

8.2.3 Install the image

```
nodeset opteron netboot
rpower opteron boot
```

8.2.3.1 Check memory usage:

```
#ssh middle "echo 3 > /proc/sys/vm/drop_caches;free -m;df -h"
              total    used    free   shared  buffers   cached
Mem:          3969     82   3887     0     0     43
-/+ buffers/cache:    38   3930
Swap:         0     0     0
Filesystem      Size  Used Avail Use% Mounted on
compute_x86_64  100M 216K 100M  1% /
none            10M   0 10M  0% /tmp
none            10M   0 10M  0% /var/tmp
```

3x smaller.

9.0 Service Node to Compute Node ssh setup

If you wish to be able to ssh from your service nodes to their compute nodes, you will have to follow these steps to add the additional required keys to the install image before the image is installed.

1. ssh to each service node
2. run `ssh-keygen -t rsa`
take default files and answer no to passcode/passphrase message
3. `cd /root/.ssh`
4. `cat id_rsa.pub >> /install/netboot/fedora8/x86_64/compute/rootimg/root/.ssh/authorized_keys`

5. `cat id_rsa.pub >>
/install/netboot/fedora8/ppc64/compute/rootimg/root/.ssh/authorized_keys`

10.0 Building image for 64K pages

On Management Node:

```
cd /opt/xcat/share/xcat/netboot/fedora  
cp compute.exlist compute.exlist.4k  
echo "./lib/modules/2.6.23.1-42.fc8/*" >>compute.exlist
```

```
wget  
http://download.fedora.redhat.com/pub/fedora/linux/releases/8/Fedora/source/SRPMS  
/kernel-2.6.23.1-42.fc8.src.rpm  
nodech rra047b nodetype.profile=iscsi  
nodeset rra047b iscsiboot  
rpower rra047b boot
```

On the blade:

```
ssh rra047b  
mkdir /install  
mount mgmt:/install /install  
yum install rpm-build redhat-rpm-config ncurses ncurses-devel kernel-devel gcc  
squashfs-tools  
rpm -Uivh kernel-2.6.23.1-42.fc8.src.rpm  
rpmbuild -bp --target ppc64 /usr/src/redhat/SPECS/kernel.spec  
cd /usr/src/redhat/BUILD/kernel-2.6.23  
cp -r linux-2.6.23.ppc64 /usr/src/  
cd /usr/src/kernels/$(uname -r)-$(uname -m)  
find . -print | cpio -dump /usr/src/linux-2.6.23.ppc64/  
cd /usr/src/linux-2.6.23.ppc64  
make mrproper  
cp configs/kernel-2.6.23.1-ppc64.config .config  
make menuconfig
```

Kernel options --->

```
[*] 64k page size
Platform support --->
[ ] Sony PS3
<exit><exit><save>
```

Edit Makefile suffix:

```
EXTRAVERSION = .1-42.fc8-64k
```

```
make -j4
make modules_install
strip vmlinux
mv vmlinux /boot/vmlinuz-2.6.23.1-42.fc8-64k
cd /lib/modules/2.6.23.1-42.fc8-64k/kernel
find . -name "*.ko" -type f -exec strip -g {} \;
#mkinitrd /boot/initrd-2.6.23.1-42.fc8-64k.img 2.6.23.1-42.fc8-64k
#rm -f /boot/vmlinuz-2.6.23.1-42.fc8 /boot/initrd-2.6.23.1-42.fc8.img
#rm -rf /lib/modules/2.6.23.1-42.fc8
```

10.1 Rebuild aufs

Rebuild aufs.so

```
rm -rf aufs
tar jxvf aufs-2-6-2008.tar.bz2
cd aufs
mv include/linux/aufs_type.h fs/aufs/
cd fs/aufs/
patch -p1 < ../../../aufs-standalone.patch
chmod +x build.sh
./build.sh 2.6.23.1-42.fc8-64k
strip -g aufs.ko
cp aufs.ko /root
```

NOTE: patch genimage

On rra047b:

```
cd /root
./genimage -i eth0 -n tg3 -o fedora8 -p compute
cd /lib/modules
cp -r 2.6.23.1-42.fc8-64k /install/netboot/fedora8/ppc64/compute/rootimg/lib/
modules/
cd /boot
cp vmlinuz-2.6.23.1-42.fc8-64k
/install/netboot/fedora8/ppc64/compute/kernel
```

10.2 Test unsquashed:**On rraa047b:**

```
cd /root
./geninitrd -i eth0 -n tg3 -o fedora8 -p compute -k 2.6.23.1-42.fc8-64k
```

On Management Node:

```
rm -f /install/netboot/fedora8/ppc64/compute/rootimg.sfs
packimage -a ppc64 -o fedora8 -p compute -m cpio
nodech rra047b nodetype.profile=compute nodetype.os=fedora8
gnodeset rra047b netboot
rpower rra047b boot
```

10.2.1 Check memory

```
# ssh left "echo 3 > /proc/sys/vm/drop_caches;free -m;df -h"
```

| | total | used | free | shared | buffers | cached |
|--------------------|-------|------|------|--------|---------|--------|
| Mem: | 4012 | 495 | 3517 | 0 | 0 | 429 |
| -/+ buffers/cache: | | 66 | 3946 | | | |
| Swap: | 0 | 0 | 0 | | | |

| Filesystem | Size | Used | Avail | Use% | Mounted on |
|---------------|------|------|-------|------|------------|
| compute_ppc64 | 2.0G | 432M | 1.6G | 22% | / |
| none | 10M | 0 | 10M | 0% | /tmp |
| none | 10M | 0 | 10M | 0% | /var/tmp |

10.3 Test squash

On rra047b:

```
cd /root
./geninitrd -i eth0 -n tg3,squashfs,aufs,loop -o fedora8 -p compute -k
2.6.23.1-42.fc8-64k -l $(expr 100 \* 1024 \* 1024)
```

On Management Node:

```
rm -f /install/netboot/fedora8/ppc64/compute/rootimg.sfs
packimage -a ppc64 -o fedora8 -p compute -m squashfs #bug, must remove
sfs first
nodech left nodetype.profile=compute nodetype.os=fedora8
nodeset left netboot
rpower left boot
```

10.3.1 Check memory

```
# ssh left "echo 3 > /proc/sys/vm/drop_caches;free -m;df -h"
```

```
total    used    free   shared  buffers   cached
Mem:      4012    127   3885     0     0     65
-/+ buffers/cache:    61   3951
Swap:      0     0     0
Filesystem      Size  Used Avail Use% Mounted on
compute_ppc64   100M  1.7M  99M   2% /
none            10M   0   10M   0% /tmp
none            10M   0   10M   0% /var/tmp
```

```
./lib/modules/* in compute.exlist:
```

10.4 Switch back to 4K pages

On rra047b

```
cd /boot
cp -f vmlinuz-2.6.23.1-42.fc8 /install/netboot/fedora8/ppc64/compute/kernel
cd /root
./geninitrd -i eth0 -n tg3 -o fedora8 -p compute
```

OR

```
./geninitrd -i eth0 -n tg3,squashfs,aufs,loop -o fedora8 -p compute -l $(expr
100 \* 1024 \* 1024)
```

From Management Node:

```
rm -f /install/netboot/fedora8/ppc64/compute/rootimg.sfs
packimage -a ppc64 -o fedora8 -p compute -m cpio
```

OR

```
packimage -a ppc64 -o fedora8 -p compute -m squashfs
nodech rra047b nodetype.profile=compute nodetype.os=fedora8
nodeset rra047b netboot
rpower rra047b boot
```

11.0 Installing OpenLDAP

11.1 Setup LDAP Server

On the management node:

1. export /home (rw) for testing
2. add a test userid : IBM

```
useradd ibm
mkdir ~ibm/.ssh
mkdir ~ibm/.pbs_spool

ssh-keygen -t rsa -q -N "" -f ~ibm/.ssh/id_rsa
cp ~ibm/.ssh/id_rsa.pub ~ibm/.ssh/authorized_keys
```

3. Create ~ibm/.ssh/config:
Add the following lines:

```
ForwardX11 yes
StrictHostKeyChecking no
FallbackToRsh no
BatchMode yes
ConnectionAttempts 5
UsePrivilegedPort no
Compression no
Cipher blowfish
UserKnownHostsFile /dev/null
CheckHostIP no
```

4. Set permissions :

```
chown -R ibm.ibm ~ibm
chmod 700 ~ibm/.ssh
chmod 600 ~ibm/.ssh/*
```

11.1.1 Install the LDAP rpms

```
yum install openldap-servers
```

The following rpms should be installed:

```
openldap-*
openldap-devel-*
```

```
openldap-clients-*
openldap-servers-*
```

11.1.2 Configure LDAP

1. `cd /etc/ldap`
2. **edit** `slapd.conf`

Put in the following information:

```
#xCAT start
```

```
#cluster.net:
```

```
suffix      "dc=cluster,dc=net"
```

```
#root access
```

```
rootdn      "cn=root,dc=cluster,dc=net"
```

```
#passwd generated with: perl -e 'print crypt("cluster","XX"),"\n"'
```

```
rootpw      {SSHA}sj0Md3HJVYLB0UY/9pou6QW7efA7dq8
```

```
# password hash algorithm
```

```
password-hash {SSHA}
```

```
# The userPassword by default can be changed by the entry owning it if they
# are authenticated. Others should not be able to see it, except the admin.
```

```
access to attrs=userPassword
```

```
    by dn="uid=admin,ou=People,dc=cluster,dc=net" write
```

```
    by anonymous auth
```

```
    by self write
```

```
    by * none
```

```
#
```

```
##password aging
```

access to attrs=shadowLastChange

by dn="uid=admin,ou=People,dc=cluster,dc=net" write

by self write

by * read

3. cp /etc/openldap/DB_CONFIG.example /var/lib/ldap/DB_CONFIG
4. start ldap
service ldap start

11.1.3 Migrate Users

cd /usr/share/openldap/migration

Edit migrate_common.ph:

```
$DEFAULT_MAIL_DOMAIN = "cluster.net";
```

```
$DEFAULT_BASE = "dc=cluster,dc=net";
```

```
$EXTENDED_SCHEMA = 1;
```

cd /usr/share/openldap/migration

```
./migrate_base.pl >/tmp/base.ldif
```

```
./migrate_passwd.pl /etc/passwd >>/tmp/base.ldif
```

```
./migrate_group.pl /etc/group >>/tmp/base.ldif
```

cd /var/lib/ldap

```
service ldap stop
```

```
slapadd -l /tmp/base.ldif
```

```
chown ldap.ldap *
```

```
service ldap start
```

11.2 Setup LDAP Client

11.2.1 Install LDAP into the image

```
yum --installroot=/install/netboot/fedora8/x86_64/compute/rootimg \  
install openldap-clients nss_ldap nfs-utils vi
```

11.2.2 Update the ldap configuration

```
cd /install/netboot/fedora8/x86_64/compute/rootimg
```

Edit /etc/ldap.conf with these changes:

```
host 11.16.0.1  
base dc=cluster,dc=net  
nss_base_passwd ou=People,dc=cluster,dc=net  
nss_base_shadow ou=People,dc=cluster,dc=net  
nss_base_group ou=Group,dc=cluster,dc=net
```

Edit etc/openldap/ldap.conf with these changes:

```
URI ldap://11.16.0.1  
BASE dc=cluster,dc=net
```

Edit etc/nsswitch with these changes

```
passwd: files ldap  
shadow: files ldap  
group: files ldap
```

Edit etc/pam.d/system-auth, change (order important!):

```
change  
account required pam_unix.so
```

to

```
account sufficient pam_ldap.so
```

```
account    required    pam_unix.so
```

Add to fstab to Mount /home for testing:

```
11.16.0.1:/home /home nfs timeo=14,intr 1 2
```

o

11.2.3 Build the image and install

Add the following rpms to the image for testing:

sunrpc,lockd,nfs,nfs_acl installed for testing (order important!):

```
cd /opt/xcat/share/xcat/netboot/fedora
```

```
./geninitrd -i eth0 -n tg3,bnx2,sunrpc,lockd,nfs,nfs_acl -o fedora8 -p compute
```

```
packimage -o fedora8 -p compute -a x86_64
```

```
nodeset rra047a netboot
```

```
rpower rra047a boot
```

12.0 Setup Hierarchical LDAP

TBD

13.0 Install Torque

13.1 Setup Torque Server

```
cd /tmp
```

```
wget
```

```
http://www.clusterresources.com/downloads/torque/torque-2.3.0.tar.gz
```

```
tar zxvf torque-2.3.0.tar.gz
```

```
cd torque-2.3.0
```

```
CFLAGS=-D__TRR ./configure \
```

```
--prefix=/opt/torque \
```

```
--exec-prefix=/opt/torque/x86_64 \
```



```
--enable-docs \  
--disable-gui \  
--with-server-home=/var/spool/pbs \  
--enable-syslog \  
--with-scp \  
--disable-rpp \  
--disable-spool  
make  
make install
```

13.2 Configure Torque

```
cd /opt/torque/x86_64/lib  
ln -s libtorque.so.2.0.0 libtorque.so.0  
echo "/opt/torque/x86_64/lib" >>/etc/ld.so.conf.d/torque.conf  
ldconfig  
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/xpbsnodes /opt/torque/x86_64/bin/  
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/pbsnodestat /opt/torque/x86_64/bin/
```

Create /etc/profile.d/torque.sh:

```
export PBS_DEFAULT=mn20  
export PATH=/opt/torque/x86_64/bin:$PATH
```

```
chmod 755 /etc/profile.d/torque.sh  
source /etc/profile.d/torque.sh
```

13.3 Define Nodes

```
cd /var/spool/pbs/server_priv  
nodels '/rr.*a' groups | sed 's/: groups:/' | sed 's/,/ /g' | sed 's/$/ np=4/' >nodes
```

13.4 Setup and Start Service

```
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/pbs /etc/init.d/  
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/pbs_mom /etc/init.d/  
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/pbs_sched /etc/init.d/  
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/pbs_server /etc/init.d/  
chkconfig --del pbs  
chkconfig --del pbs_mom  
chkconfig --del pbs_sched  
chkconfig --level 345 pbs_server on  
service pbs_server start
```

13.5 Install pbstop

```
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/pbstop /opt/torque/x86_64/bin/  
chmod 755 /opt/torque/x86_64/bin/pbstop
```

13.6 Install Perl Curses for PBS top

```
cd /tmp  
tar zxvf /opt/xcat/share/xcat/netboot/add-on/torque/Curses-1.23.tgz  
cd Curses-1.23  
perl Makefile.PL  
make  
make install
```

13.7 Create a Torque default queue

```
echo "create queue dque  
set queue dque queue_type = Execution  
set queue dque enabled = True
```

```
set queue dque started = True
set server scheduling = True
set server default_queue = dque
set server log_events = 127
set server mail_from = adm
set server query_other_jobs = True
set server resources_default.walltime = 00:01:00
set server scheduler_iteration = 60
set server node_pack = False
s s keep_completed=300" | qmgr
```

13.8 Setup Torque Client (x86_64 only)

13.8.1 Install Torque

```
cd /opt/xcat/share/xcat/netboot/add-on/torque
./add_torque /install/netboot/fedora8/x86_64/compute/rootimg mn20 /opt/torque
x86_64 local
```

13.8.2 Configure Torque

13.8.2.1 Setup Access

```
cd /install/netboot/fedora8/x86_64/compute/rootimg/etc/security
echo "-:ALL EXCEPT root:ALL" >>access.conf
cp access.conf access.conf.BOOT
```

```
cd /install/netboot/fedora8/x86_64/compute/rootimg/etc/pam.d
```

```
edit system-auth
```

replace:

```
account sufficient pam_ldap.so
```

```
account required pam_unix.so
```

with:

```
account required pam_access.so
```

```
account sufficient pam_ldap.so
```

```
account required pam_unix.so
```

13.8.2.2 Setup node to node ssh for root

This is needed for cleanup

```
cp /root/.ssh/* /install/netboot/fedora8/x86_64/compute/rootimg/root/.ssh/
```

```
cd /install/netboot/fedora8/x86_64/compute/rootimg/root/.ssh/
```

```
rm known_hosts
```

Setup the config file:

```
echo "StrictHostKeyChecking no
```

```
FallBackToRsh no
```

```
BatchMode yes
```

```
ConnectionAttempts 5
```

```
UsePrivilegedPort no
```

```
Compression no
```

```
Cipher blowfish
```

```
CheckHostIP no" >config
```

13.8.3 Pack and Install image

```
packimage -o fedora8 -p compute -a x86_64
```

```
nodeset opteron netboot
```

```
rpower opteron boot
```

14.0 Setup Moab

14.1 Install Moab

```
cd /tmp
wget http://www.clusterresources.com/downloads/mwm/moab-5.2.1-linux-x86_64-
torque.tar.gz
tar zxvf /tmp/moab-5.2.1-linux-x86_64-torque.tar.gz
cd moab-5.2.1
./configure --prefix=/opt/moab
make install
```

14.2 Configure Moab

```
mkdir -p /var/spool/moab/log
mkdir -p /var/spool/moab/stats
```

Create /etc/profile.d/moab.sh:

```
export PATH=/opt/moab/bin:$PATH
```

```
chmod 755 /etc/profile.d/moab.sh
source /etc/profile.d/moab.sh
```

Edit moab.cfg,

change:

```
RMCFG[mn20] TYPE=NONE
```

to:

```
RMCFG[mn20] TYPE=pbs
```

Append to moab.cfg :

```
NODEAVAILABILITYPOLICY DEDICATED:SWAP
```

```
JOBNODEMATCHPOLICY    EXACTNODE
NODEACCESSPOLICY      SINGLEJOB
NODEMAXLOAD           .5
JOBMAXSTARTTIME       00:05:00
DEFERTIME              0
JOBMAXOVERRUN         0
LOGDIR                 /var/spool/moab/log
LOGFILEMAXSIZE        10000000
LOGFILEROLLDEPTH      10
STATDIR                /var/spool/moab/stats
```

14.2.1 Start Moab

```
cp -f /opt/xcat/share/xcat/netboot/add-on/torque/moab /etc/init.d/
chkconfig --level 345 moab on
service moab start
```

15.0References

XCAT2.0 Beta Cookbook - <http://xcat.sourceforge.net/xCAT2.0.pdf>