# XCAT 2 on AIX

# Creating an xCAT Management Node

07/05/2010, 09:29:14 AM


1.0 Overiew ..... 1
2.0 Installing xCAT and prerequisite Software ..... 1
2.1 Set up an AIX system to use as an xCAT Management Node ..... 2
2.2 Install AIX prerequisite software ..... 2
2.2.1 openssl and openssh ..... 2
2.2.2 expect, tk, and tcl ..... 2
2.2.3 devices.tmiscw (Optional) ..... 3
2.3 Create a new volume group for your /install directory (optional) ..... 3
2.4 Download and install the prerequisite Open Source Software (OSS) ..... 4
2.5 Download and install the xCAT software ..... 4
2.6 Verify the xCAT installation ..... 5
3.0 Additional configuration of the management node ..... 6
3.1 Cluster network configuration notes ..... 6
3.2 Choose the shell to use in the cluster (optional) ..... 6
3.3 Configuring name resolution (optional) ..... 6
3.3.1 Add cluster nodes to the /etc/hosts file ..... 7
3.3.2 Set up a DNS nameserver ..... 8
3.4 Syslog setup ..... 9
3.5 Add cluster resolv.conf file (optional) ..... 10
3.6 Set cluster root password (optional) ..... 10
3.7 Set up NTP (optional) ..... 10
3.8 Increase file size limit ..... 11
3.9 Check the policy definitions ..... 11
3.10 Check system services ..... 11
4.0 xCAT on AIX documentation ..... 12
4.1 Terminology ..... 12
4.2 Installing AIX standalone nodes (using NIM rte method) ..... 13
4.3 Booting AIX diskless nodes (using stateless method) ..... 13
4.4 Cloning AIX nodes (install using AIX mksysb image) ..... 13
4.5 Using xCAT Service Nodes with AIX ..... 13
4.6 Updating AIX cluster nodes ..... 13
5.0 References ..... 13


## 1.0 Overiew

This document describes how to install and configure an xCAT on AIX management node.

## 2.0 Installing xCAT and prerequisite Software

## 2.1 Set up an AIX system to use as an xCAT Management Node

- Follow AIX documentation and procedures to install and configure the base AIX operating system. (Typically by using the product media.)
- Apply the latest AIX software updates and fixes as needed.
- Make sure the OS version installed on the management node is greater than or equal to the OS versions you wish to install on the cluster nodes.

## 2.2 Install AIX prerequisite software

To install the additional AIX software you have a choice of several different interfaces provided by AIX. Perhaps the easiest method is to use the SMIT (or "smitty") interface but you could also use the AIX **geninstall, installp, or rpm** commands if you like. Refer to the AIX documentation if you are not familiar with this support. (http://www-03.ibm.com/servers/aix/library/index.html)

**Important Note**
Since these are **installp** file sets you should run **/usr/sbin/updtvpkg** after installing to make sure that the RPM reflection of what was installed by **installp** is updated. This makes it possible for RPM packages with a dependencies to recognize that the dependency is satisfied.

***updtvpkg***

### 2.2.1 openssl and openssh

The openssl and openssh installp filesets are now available on AIX product media. (Starting in AIX 6.1.3.)

You can check to see if these are installed by running the "**lslpp**" command. For example:

***lslpp -l | grep open***

If they are not installed then use the AIX product media and standard AIX tools to install them.

### 2.2.2 expect, tk, and tcl

This software is now shipped with AIX product media. (Starting in AIX 6.1.2.)

They are normally installed with AIX but in some cases you will have to install them manually from the AIX media.

Check if they are installed and install them if needed.

If they are not available on the AIX media then you can get them from the "AIX Toolbox for Linux Applications" (http://www-03.ibm.com/systems/power/software/aix/linux/toolbox/alpha.html)

### 2.2.3 devices.tmiscw (Optional)

If you plan to be using AIX diskless nodes and you wish to set up system dump support for those nodes then you will need the devices.tmiscw software installed on your management node. This software is available on the AIX Expansion Pack.

Install the software using standard AIX interfaces.

**Note:** The xCAT diskless dump support is available in xCAT version 2.5 and beyond. You will also need AIX 6.1.6 or greater for full support.

## *2.3 Create a new volume group for your /install directory (optional)*

By default xCAT uses the /install directory to store various xCAT and NIM resources. XCAT will create /install as a subdirectory of the / (root) file system. In some cases /install may not contain enough space for your intended use.

To avoid this problem you could create a separate file system called /install on the management server to store the files that are to be used with xCAT and NIM. The size of this file system depends on your particular cluster.

The largest files that will be stored in /install subdirectories will be the NIM resources required for installing AIX nodes. The space required for a unique set of AIX operating system installation resources is approximately 2.0 GB. If you will need to manage several levels of OS images you should plan on at least 2G for each.

You can create the /install file system as part of the rootvg or in its own volume group. The following examples illustrate how to create the /install file system using the root volume group. To create a 5 GB file system called /install you could issue the AIX **crfs** command:

```
crfs -v jfs2 -g rootvg -m /install -a size=5G -A yes
```

After you have created /install, you must mount it, as follows:

```
mount  /install
```

**Note:** You can use the AIX SMIT interfaces to create new volume groups and file systems etc. For example, to create a new file system you could use the SMIT fastpath ("crfs") to go directly to the correct SMIT panel. (Just type "*smit crfs".)*

## 2.4 Download and install the prerequisite Open Source Software (OSS)

- Download the latest dep-aix-*.tar.gz tar file from http://xcat.sourceforge.net/#download and copy it to a convenient location on your xCAT management node.
- Unwrap the tar file. For example:

  ***gunzip dep-aix-*.tar.gz***
  ***tar –xvf dep-aix-*.tar***

- Read the README file.
- Run the **instoss** script (contained in the tar file) to install the OSS packages. Please make sure the /opt and the other file systems have enough disk space to install these OSS packages before running the **instoss** script.

**Note:** The expect, tk and tcl rpms are no longer shipped by xCAT. They are now shipped with AIX.

**Note #2**: For easier downloading without a web browser, you may want to download and install the **wget** tool from the AIX Toolkit for Linux.

## 2.5 Download and install the xCAT software.

- Download the latest xCAT for AIX tar file from http://xcat.sourceforge.net/#download and copy it to a convenient location on your xCAT management node.
- Unwrap the xCAT tar file. For example,

  ***gunzip core-aix-*.tar.gz***
  ***tar –xvf core-aix-*.tar***

- Run the **instxcat** script (contained in the tar file) to install the xCAT software. (For xCAT 2.3.2 and above, the **instxcat** script and all the RPMs are located in the xcat-core subdirectory.) The post processing provided by the xCAT packages will perform some basic xCAT configuration. (This includes initializing the SQLite database and starting **xcatd** daemon processes.) Note: xCAT software packages will install about 200MB files into /opt directory, make sure the /opt directory has enough disk space before running **instxcat** script.
- Execute the system profile file to set the xCAT paths. This file was updated during the xCAT post install processing. (“. */etc/profile”)*. ( **Note**: Make sure you don't have a .profile file that overwrites the “PATH” environment variables.)

## 2.6 Verify the xCAT installation.

- Run the "*lsdef –h*" to check if the xCAT daemon is working. (If you get a correct response then you should be Ok. )

- Check to see if the initial xCAT definitions have been created. For example, you can run "*lsdef -t site -l*" to get a listing of the default site definition. You should see output similar to the following.

```
------------------------------------------------------------
Setting the name of the site definition to 'clustersite'.

Object name: clustersite
   domain=abc.foo.com
   installdir=/install
   tftpdir=/tftpboot
   master=7.104.46.27
   useSSHonAIX=yes
   xcatdport=3001
   xcatiport=3002
```

**Important**: The "domain" and "master" values are set automatically by xCAT when it is installed. To do this xCAT looks at the primary hostname of the management node.

For the "domain" attribute, if the management node hostname was set to a short hostname then the domain attribute would not be set by default. It is also possible that the domain would be set to a value other than the domain that is used for the cluster nodes. In either case you must manually set the domain value to the network domain that will be used for the cluster nodes. You can use the xCAT **chdef** command to modify the domain attribute of the cluster site definition.

For example:

***chdef -t site domain=mycluster.com***

The "master" attribute must be set to the hostname of the xCAT management node, as known by the nodes.

For example:

***chdef -t site master=xcatmn***

# 3.0 Additional configuration of the management node

## *3.1 Cluster network configuration notes*

- The cluster network topology, naming conventions etc. should be carefully planned before beginning the cluster node deployment.
- XCAT requires an Ethernet network for installing and managing cluster nodes.
- Cluster nodes may be on different subnets.
- The cluster nodes must all have unique short host names to use in the xCAT node definitions.
- All cluster nodes must use the same domain name. The domain attribute must be set in the cluster site definition.
- The management node interfaces that will be used to manage the nodes should be configured before starting the xCAT deployment process.
- XCAT network definitions will have to be created for each unique subnet used in the cluster. (This will be described in one of the install documents listed below.)
- If you will be using the xCAT management node or a service node as a gateway remember to set "ipforwarding" to "1".

## *3.2 Choose the shell to use in the cluster (optional)*

By default the xCAT support will automatically set up **ssh** on all AIX cluster nodes. If you wish to use r**sh** you should modify the cluster site definition. To use r**sh** you would have to set the "useSSHonAIX=no". You can also specify a path for the **ssh** and **scp** commands by setting the "**rsh**" and "**rcp**" . If not set the default path would be "/usr/bin/ssh" and "/usr/bin/scp".

You will also have to make sure that the **openssl** an **openssh** software is installed on your nodes. This is covered in the cluster node installation documents listed below.

To change the shell you must change the value of the *useSSHonAIX* attribute in the cluster site definition. For example:

***chdef -t site useSSHonAIX=no***

**Note**: If, at some future point, you wish to check which shell is being used you can run **xdsh** to a node with the "-T" (trace) option. For example:

*xdsh node01 -v -T date*

Note: The default shell for xCAT 2.3 and beyond is **ssh**. In earlier versions of xCAT the default was **rsh**.

## *3.3 Configuring name resolution (optional)*

Name resolution is required by xCAT. You can use a simple /etc/hosts mechanism or you can optionally set up a DNS name server. In either case you must start by setting up the /etc/hosts file.

**If you do not set up DNS you may need to distribute new versions of the /etc/hosts file to all the cluster nodes whenever you add new nodes to the cluster.**

### 3.3.1 Add cluster nodes to the /etc/hosts file

There are several ways to get entries for all the cluster nodes in the /etc/hosts file.

These include:

- Manually adding the entries.
- Running a custom script that uses some cluster naming convention to automate the adding of the node entries. (User-provided.)
- Using the xCAT **makehosts** command after the XCAT node definitions have been created.

If you are dealing with a large number of nodes this task can be quite tedious. The xCAT **makehosts** option may be useful in some cases. This process uses a regular expression to automatically determine the IP addresses and hostnames for a set of nodes. To use this method you must decide on appropriate naming conventions and IP address ranges for your nodes. This process may seem a bit complicated but once you get things set up it can save time and add structure to your cluster.

If you choose to use this process you will have to come back to this section after you have created the xCAT node definitions later in this process. You should read through this now and decide on naming conventions etc. for when you create your xCAT node definitions.

The basic process is:

- Decide on a node naming convention such that the node IP & long hostname can be determined from the node name.
- Include all the nodes in a node "group" definition.
- Set the group "ip" and "hostnames" attribute to a regular expression that can be used to derive the node IP and hostname.
- Run the **makehosts** command to add all the node information to the /etc/hosts file.

As an example, suppose we decide on a node naming convention that includes the hardware frame number, the CEC number and the partition number. (Say "clstrf01c01p01" etc.) Also, lets say that the IP addresses would look something like "100.1.1.1" where the second number is the frame number, the third is the CEC number and the forth is the partition number.

With this example we can define a regular expression that, given a node name, could be used to derive a corresponding IP address and long hostname.

To have this regular expression applied to each node you can make use of the xCAT node group support. Let's say that all your cluster nodes belong to the group "compute". I can add the following values to the "compute" group definition.

***chdef -t group -o compute ip='|clstrf(\d+)c(\d+)p(\d+)|10.($1+0).($2+0).($3+0)|' hostnames='|(.*)|($1).cluster.com|'***

This basically says that for any node in the "compute" group the "ip" can be derived by the regular expression *'|clstrf(\d+)c(\d+)p(\d+)|10.($1+0).($2+0).($3+0)|'* , and the hostname can be derived from the expression *|(.*)|($1).mycluster.com|'*.

So let's say that you have defined all your nodes using the xCAT support such as **rscan** or **mkvm** using the naming convention mentioned above. Now you could display the node definition as follows:

*lsdef -l clstrf01c02p03*

Since this node belongs to the "compute" group, when I display the definition it will use the regular expressions to derive the "ip" and "hostnames" values.

The output might look something like the following:

*Object name: clstrf01c02p03*
*cons=hmc*
*groups=lpar,all,compute*
*hcp=clstrhmc01*
*hostnames=clstrf01c02p03.mycluster.com*
*id=1*
*ip=10.1.2.3*
*mac=001a64f9c009*
*mgt=hmc*
*nodetype=lpar,osi*
*os=AIX*
*parent=clstrf1fsp01-9125-F2A-SN024C332*
*postscripts=myscript*
*profile=MYimg*

Now that all the nodes have an "ip" and "hostnames" value you can run the xCAT **makehosts** command to update /etc/hosts.

***makehosts compute -l***

### 3.3.2 Set up a DNS nameserver

To set up the management node as the DNS name server you must set the "domain", "nameservers" and "forwarders" attributes in the xCAT "site" definition.

For example, if the cluster domain is “mycluster.com”, the IP address of the management node is “*100.0.0.41*” and the site DNS servers are “9.14.8.1,9.14.8.2” then you would run the following command.

***chdef -t site domain= mycluster.com nameservers= 100.0.0.41 forwarders= 9.14.8.1,9.14.8.2***

Edit “/etc/resolv.conf” to contain the cluster domain and nameserver. For example:

*search mycluster.com*
*nameserver 100..0.0.41*

Create xCAT network definitions for each of the cluster networks. (Your network and mask value need to be defined for **makedns** to be able to set up the correct ip range for the management node to serve.)

You will need a name for the network and values for the following attributes.

**net** The network address.
**mask** The network mask.
**gateway** The network gateway.

You can use the xCAT **makenetworks** command to gather cluster network information and create xCAT network definitions. See the **makenetworks** man page for details. (This feature is available in xCAT 2.3 and beyond.)

You can also use the xCAT **mkdef** command to define the network.

For example:

***mkdef -t network -o net1 net=9.114.113.224 mask=255.255.255.224 gateway=9.114.113.254***

Run **makedns** to create the /etc/named.conf file and populate the /var/named directory with resolution files.

***makedns***

Start DNS:

***startsrc -s named***

## *3.4 Syslog setup*

xCAT will automatically set up **syslog** on the management node and the cluster nodes when they are deployed (installed or booted). When **syslog** is set up on the nodes it will be configured to forward the logs to the management node.

If you do not wish to have **syslog** set up on the nodes you must remove the "syslog" script from the "xcatdefaults" entry in the xCAT "postscripts" table. You can change the "*xcatdefaults*" setting by using the xCAT **chtab** or **tabedit** command.

## 3.5 Add cluster resolv.conf file (optional)

The xCAT deployment code will automatically handle the creation of an /etc/resolv.conf file on all the cluster nodes. If you want xCAT to handle this you should make sure the "domain" and "nameservers" attributes of the "site" definition are set.

For example:

***chdef -t site -o clustersite domain=mycluster.com nameservers= 100.240.0.1***

## 3.6 Set cluster root password (optional)

You can have xCAT create an initial root password for the cluster nodes when they are deployed. To do this you must modify the xCAT "passwd" table.

You can use the **tabedit** command to add and entry to this table. For example:

***tabedit passwd***

You will need an entry with a *"key"* set to *"system"*, a *"username"* set to *"root"* and the *"password"* attribute set to whatever string you want.

In xCAT version 2.5 and beyond you may add an encrypted password to the table. If the password is encrypted you must also set the "cryptmethod" attribute so that the password can be set correctly on the nodes.

You can change the passwords on the nodes at any time using **xdsh** and the AIX **chpasswd** command.

For example:

*xdsh node01 'echo "root:mypw" | chpasswd -c'*

## 3.7 Set up NTP (optional)

To enable the NTP services on the cluster, first configure NTP on the management node and start **ntpd**.

Next set the "*ntpservers"* attribute in the site table. Whatever time servers are listed in this attribute will be used by all the nodes that boot directly from the management node.

If your nodes have access to the internet you can use the global servers:

*chdef -t site ntpservers= 0.north-america.pool.ntp.org,*
*1.northamerica.pool.ntp.org,2.north-*
*america.pool.ntp.org,3.northamerica.pool.ntp.org*

If the nodes do not have a connection to the internet (or you just want them to get their time from the management node for another reason), you can use your management node as the NTP server. For example, if the name of you management node is "myMN" then you could run the following command.

***chdef -t site ntpservers= myMN***

## 3.8 Increase file size limit

Some of the AIX/NIM resources that are used to install nodes are quite large (1-2G) so it may be necessary to increase the file size limit.

For example, to set the file size limit to "unlimited" for the user "root" you could run the following command.

***/usr/bin/chuser fsize=-1 root***

## 3.9 Check the policy definitions.

When the xCAT software was installed it created several policy definitions. To list the definitions you can run:

*lsdef -t policy -l*

You may need to add additional policy definitions. For example, you will need a policy for the hostname that was used when xCAT was installed. To find out what this was you can run:

*openssl x509 -text -in /etc/xcat/cert/server-cert.pem -noout|grep Subject:*

So, for example, if the hostnmae is "myMN.foo.bar" then you can create a policy definition with the following command. (The policy names are numbers, just pick a number that is not yet used.)

*mkdef -t policy -o 8 name= myMN.foo.bar rule=allow*

## 3.10 Check system services

- **inted**

  inetd includes services such as telnet, ftp, bootp/dhcp, and others. **Edit the /etc/inetd.conf file to turn on all services that are needed.** Ftp and bootp/dhcp are required for System p node installations. Stop and restart the **inetd** service after any changes:

  *stopsrc -s inetd*
  *startsrc -s inetd*

- **NFS**

  NFS is required for all NIM installs. Ensure the NFS daemons are running:
  *lssrc -g nfs*
  If any NFS services are inoperative, you can stop and restart the entire group of services:
  *stopsrc -g nfs*
  *startsrc -g nfs*
  There are other system services that NFS depends on such as inetd, portmap, biod, and others.

- **TFTP**
  To check if the TFTP daemonis running.
  *lssrc -a | grep tftpd*

  To stop and start tftp daeon.
  *stopsrc -s tftpd*
  *startsrc -s tftpd*

# 4.0 xCAT on AIX documentation

## *4.1 Terminology*

Some basic terminology.

- **standalone –** An AIX node that has it's operating system installed on a local disk.
- **rte** install - A network installation method supported by NIM that uses a NIM lpp_source resource to install a node.
- **mksysb** install - A network installation method supported by NIM that uses a system backup of one node (mksysb image) to install other cluster nodes.
- **diskful** – A node that has its operating sytem installed on a local disk. (A standalone node.)
- **diskless** - No local disk. For AIX nodes the operating system is mounted.
- **stateful** – A node that maintains its state after it has been shut down and rebooted. For AIX diskless nodes this means that each node has its own NIM "root" resource that it can use to store node-specific information. Each node mounts its own root directory.
- **stateless** – A node that does NOT maintain its state after it has been shut down and rebooted. For AIX diskless nodes this means that the nodes all use the same NIM "shared_root" resource. Each node mounts the same root directory. Anything that is written to the local root directory is redirected to memory and is lost when the node is shut down.
- **statelite** - An AIX diskless stateless node that also has a small amount of persistent files and/or directories. The persistent files and/or directories are

mounted on the nodes. This support is available for AIX nodes in xCAT version 2.5 and beyond.

### 4.2 Installing AIX standalone nodes (using NIM rte method)

http://xcat.svn.sourceforge.net/viewvc/xcat/xcat-core/trunk/xCAT-client/share/doc/xCAT2onAIXinstall.pdf

### 4.3 Booting AIX diskless nodes (using stateless method)

http://xcat.svn.sourceforge.net/viewvc/xcat/xcat-core/trunk/xCAT-client/share/doc/xCAT2onAIXDiskless.pdf

### 4.4 Cloning AIX nodes (install using AIX mksysb image)

http://xcat.svn.sourceforge.net/viewvc/xcat/xcat-core/trunk/xCAT-client/share/doc/xCAT2onAIXmksysb.pdf

### 4.5 Using xCAT Service Nodes with AIX

http://xcat.svn.sourceforge.net/viewvc/xcat/xcat-core/trunk/xCAT-client/share/doc/xCAT2onAIXServiceNodes.pdf

### 4.6 Updating AIX cluster nodes

http://xcat.svn.sourceforge.net/viewvc/xcat/xcat-core/trunk/xCAT-client/share/doc/xCAT2onAIXUpdates.pdf

## 5.0 References

- xCAT man pages: http://xcat.sf.net/man1/xcat.1.html
- xCAT DB table descriptions: http://xcat.sf.net/man5/xcatdb.5.html
- xCAT mailing list: http://xcat.org/mailman/listinfo/xcat-user
- xCAT bugs: https://sourceforge.net/tracker/?group_id=208749&atid=1006945
- xCAT feature requests: https://sourceforge.net/tracker/?group_id=208749&atid=1006948
- xCAT wiki: http://xcat.wiki.sourceforge.net/