

xCAT 2 on AIX (Beta version)

06/19/2008

1.0 Release Description	1
1.1.1 xCAT for AIX Support Summary	1
1.1.2 Licensing	2
2.0 Overview of xCAT support for AIX	2
2.1 xCAT Commands	2
2.2 Database tables	2
2.3 Object definitions	3
2.4 Syslog setup	4
2.5 Remote shell setup	4
2.5.1 Setting up OpenSSH	4
2.6 NTP setup	5
2.7 Customization script support	5
2.8 Operating system image management	6
2.9 Hardware discovery	6
2.10 LPAR management	7
2.11 Hardware control	7
2.12 Remote console support	8
2.13 xCAT Notification and Monitoring Plug-in Infrastructure	8
3.0 Installing xCAT and prerequisite Software	8
4.0 Deploying AIX nodes	11
4.1 Installing AIX nodes (using standard NIM rte method)	11
4.2 Booting AIX diskless nodes	11
4.3 Cloning AIX nodes using an AIX mkysyb image	11
5.0 References	11

1.0 Release Description

xCAT (Extreme Cluster Administration Tool) is a toolkit that provides support for the deployment and administration of cluster environments.

Earlier versions of xCAT have been used to deploy and manage many high end Linux clusters. The new version 2 of xCAT is a complete rewrite of xCAT that includes many architectural changes and functional enhancements.

This document describes the features provided by the beta release of xCAT support for an AIX cluster. It focuses on xCAT support unique to AIX systems.

1.1.1 xCAT for AIX Support Summary

The support provided by xCAT for AIX clusters includes the following:

- Deploying AIX diskless and standalone nodes.

- Automation of NIM setup.
- Support for AIX 5.3 and higher releases.
- Operating system image management.
- User-provided customization scripts.
- Hardware control commands for discovering hardware, gathering MAC addresses, VPD, and environmentals, power control, initiating a network boot, and LPAR creation/deletion.
- LPAR configuration management.
- Support for POWER 5 and POWER 6 hardware.
- Remote console support.
- Parallel remote shell and remote copy commands.
- Monitoring plug-in infrastructure
- RMC monitoring plug-in.
- Notification infrastructure which lets users monitor xCAT database table changes.
- Predefined conditions, responses and sensors.
- Automatic setup for syslog, remote shell, and ntp for both the xCAT management node and the cluster nodes.
- xCAT cluster data stored in a relational database.
- Commands to manipulate the xCAT database tables directly.
- Commands to manage xCAT data object definitions.
- Documentation and “**man**” pages.

1.1.2 Licensing

xCAT 2.0 is open source software with an EPL license. For license information visit:

<http://www.opensource.org/licenses/eclipse-1.0.php>

2.0 Overview of xCAT support for AIX

2.1 xCAT Commands

See <http://xcat.sf.net/man1/xcat.1.html> for a description of the supported xCAT commands.

2.2 Database tables

The xCAT data that is used to manage a cluster is contained in a relational database. Different types of data are stored in different tables. You can manage this information directly using a set of table oriented commands provided by xCAT.

See <http://xcat.sf.net/man5/xcatdb.5.html> for a description of the xCAT tables.

To manage these tables directly, xCAT provides the **chtab**, **tabdump**, **tabrestore**, and **tabedit** commands.

2.3 Object definitions

In addition to managing the database tables directly, xCAT also supports the concept of data object definitions. Data objects are abstractions of the data that is stored in the xCAT database. This support provides a conceptually simpler implementation for managing cluster data, (especially data associated with a specific cluster node). It is also more consistent with other IBM systems management products. The attributes and values defined in the data object definitions will still be stored in the database tables defined for xCAT. These data object definitions should not limit experienced xCAT customers from managing the specific tables directly, if they so desire. A new set of commands is provided to support the object definitions. These commands will automatically handle the storage in and retrieval from the correct tables.

The following data object types are currently supported.

- **site** - Cluster-wide information. All the data is stored in the *site* table.
- **node** - Information for a specific cluster node. The data for a node is stored in multiple tables in the database. The commands that are provided to manage these definitions automatically figure out which attributes are stored in which table. It is therefore not necessary to keep track of a large number of table names and attribute locations.
- **network** - A description of a unique network. This data is stored in the *networks* table.
- **monitoring** - A description of a monitoring plugin. This data is stored in the *monitoring* table.
- **notification** - Defines the Perl modules and commands that will get called for changes in certain xCAT database tables. The data is stored in the *notification* table.
- **group** - Defines a set of nodes. A group definition can be used as the target set of nodes for a specific xCAT operation. It can also be used to define node attributes that are applied to all group members. The group data is stored in multiple tables in the database.
- **osimage** - Defines a unique operating system image and related resources that are required for xCAT to deploy a cluster node.
- **policy** - Controls who has authority to run specific xCAT operations.

There are four basic xCAT commands that may be used to manage any of the data object definitions.

- **mkdef** - Make data object definitions.
- **chdef** - Change data object definitions.
- **lsdef** - List data object definitions.
- **rmdef** - Remove data object definitions.

In addition to the standard command line input and output the **mkdef**, **chdef**, and **lsdef** commands support the use of a stanza file format for the input and output of

information. Input to a command can be read from a stanza file and the output of a command can be written to a stanza file. A stanza file contains one or more stanzas that provide information for individual object definitions. For example:

1. To create a set of definitions using information contained in a stanza file.
`cat mystanzafile | mkdef -z`
2. To write all node definitions to a stanza file.
`lsdef -t node -l -z > nodestanzafile`

The stanza file support also provides an easy way to backup and restore the cluster data.

For more information on the use of stanza files see the **xcatstanzafile man** page.

2.4 Syslog setup

xCAT will automatically set up syslog on the management node and the cluster nodes when they are deployed. 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.

2.5 Remote shell setup

If you wish to use the **xdsh**, **xdcp**, or **xdshbak** commands you will have to configure either **ssh** or **rsh**. The default for xdsh, xdcp is rsh. To change to use ssh, then follow the below instructions. **psh** only supports **ssh**.

2.5.1 Setting up OpenSSH

To setup the remote shell during the node installation:

- Update the site table.
rsh is setup by default. To request the setup of ssh instead of rsh, set the useSSHonAIX attribute in the site table to “yes” or “1”.
- Update the poscripts table

Add aixremoteshell to the list of xcatdefaults postscripts to be run in the postscripts table. It should be added after the syslog postscript. aixremoteshell will setup either rsh or ssh depending on the setting of the useSSHonAIX attribute in the site table.

The aixremoteshell script will setup the capability to ssh from the Management Node to the nodes as root without being prompted for a password. This is required for using xdsh and xdcp with ssh.

If you also require root to be able to ssh from the nodes to the Management node and between nodes, then after the install and ssh is setup by aixremoteshell, run the following command to all, or any group of nodes that needs this capability. This will transfer the private ssh key to the nodes.

Note: there are security risks with doing this, in that if root on one node is compromised, root can login to any node in the system without a password.

```
xdcp <nodegroup> /.ssh/id_rsa /.ssh/id_rsa
```

- Install OpenSSH & OpenSSL software.

Because **OpenSSH** and its prerequisite **OpenSSL** are not installed with base AIX, you will have to install this additional software on the xCAT management node and the AIX cluster nodes.

The latest versions of the software are available from the AIX Expansion Pack. The software is also available to download from the following sites.

OpenSSL

<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=aixbp>

OpenSSH

<http://sourceforge.net/projects/openssh-aix>

This software could be installed on the nodes during the initial node installation using NIM or you could just copy the packages to the node and use the standard installp command. See the AIX NIM documentation for details.

2.6 NTP setup

To have xCAT setup ntp on the nodes during the install:

- Update the postscripts table

Add setupntp to the list of xcatdefaults postscripts to be run in the postscripts table. setupntp should be added after syslog.

2.7 Customization script support

xCAT supports the automatic running of user-provided customization scripts on the nodes when they are deployed.

To have your script run on the nodes:

1. Put a copy of your script in `/install/postscripts` on the xCAT management node. (Make sure it is executable.)
2. Set the “postscripts” attribute of the node or group definition to include a comma separated list of the scripts that you want to be executed on the nodes. For example, if you want to have your two scripts called “foo” and “bar” run on node “*node01*” you could use the **chdef** command as follows.

```
chdef -t node -o node01 postscripts=foo,bar
```

The order of the scripts in the list determines the order in which they will be run.

xCAT also runs some basic scripts to do default node configuration. You can see what default scripts xCAT will run by looking at the “xcatdefaults” entry in the xCAT “postscripts” database table. (I.e. Run “`tabdump postscripts`”). You can change the “xcatdefaults” setting by using the xCAT **chtab** or **tabedit** command.

2.8 Operating system image management

xCAT for AIX includes support for managing different “flavors” of operating system images that may be used to deploy the cluster nodes. Each unique “flavor” is described in an xCAT “*osimage*” object definition. An *osimage* definition includes the information that xCAT needs to deploy a node.

The specific contents of an *osimage* definition will vary depending on the type of OS image and how it will be deployed.

When dealing with AIX systems the xCAT *osimage* definition is typically created automatically when using xCAT commands (ex. **mknimimage**) to create the AIX/NIM resources that are needed to deploy the nodes. The definition may be modified, listed, and removed as needed using the xCAT **mkdef**, **chdef**, **lsdef**, and **rmdef** commands.

An *osimage* may be associated with multiple nodes. The “profile” attribute value for a node is set to the name of the *osimage* that will be used to deploy the node.

The *osimage* definition is used by the relevant xCAT commands (ex. **mkdsklnode**, **nimnodeset**) to automatically determine what setup must be done to deploy a node with the specified OS image.

2.9 Hardware discovery

There are several xCAT commands that can be used to gather information about cluster hardware.

- **rinv** - Retrieves hardware configuration information for a single or range of nodes and groups.
- **rvitals** - Retrieves hardware vital information for a single or range of nodes and groups.
- **lsslp** - Queries selected networked services information within the same subnet. If the HMC/IVM that you are interested in discovering is on the same subnet as your Management Node, you can run the **lsslp** to discover and add this hardware to the xCAT database.

Note that the dependent programs **slp_query** and **libslp_client.so** are compiled modules required to perform SLP broadcasts. These modules can be obtained by posting a request to the xCAT mailing list (please specify the target O/S in the request).

- **rscan** - Collects node information from one or more hardware control points.
- **getmacs** – Gathers adapter MAC information from cluster nodes. For example:

2.10 LPAR management

xCAT commands can be used to help manage the configuration of POWER 5 and POWER 6 logical partitions.

You can:

- Create new partition(s) with the same profile/resources as an existing partition. (See the **mkvm** command.)
- Duplicate all the partitions associated with one CEC in another CEC (with the same HMC). (See the **mkvm** command.)
- List partition profiles. (See the **lsvm** command.)
- Modify partition profiles. (See the **chvm** command.)
- Remove partitions. (See the **rmvm** command.)
- Automatically switch to a new partition profile when network booting a node using **rnetboot**.

Sample scenarios: !!!! need more info - details!!!!!!

1. Create 40 partitions all using the same profile.
 - Create an initial partition using the HMC interfaces.

- Use mkvm to create 40 partitions using this profile.
2. Switch a node to a new partition profile.
 - Create a new partition profile using the HMC interfaces
 - Set the “pprofile” attribute of the node definition to the name of the new partition profile.
 - Do the required xCAT node setup and initiate a network boot of the node.
 - 3.

2.11 Hardware control

The following xCAT :q

2. commands can be used to control cluster hardware.
 - **rnetboot** – Initiate a network boot request on one or more cluster nodes.
 - **rpower** – Boots, resets, powers on and off, and queries node hardware, and devices.

2.12 Remote console support

You can open a remote console using the xCAT **rcons** command. This command requires that you have conserver installed and configured. To configure conserver, run the xCAT command:

```
makeconservercf
```

Kill the conserver daemon if it is running, and restart it:

```
conserver &
```

(You may need to add /opt/freeware/bin and /opt/freeware/sbin to your PATH first).

For example, to start a console for *node01*.

```
rcons node01
```

2.13 xCAT Notification and Monitoring Plug-in Infrastructure

- add link to Ling's doc????????????????????

3.0 Installing xCAT and prerequisite Software

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.)

You may also choose to use NIM from another cluster to initially install your MN. Just remember to update your resources appropriately (especially your resolv.conf resource),

- Apply the latest software updates and fixes.
- Install the latest versions of OpenSSL & OpenSSH from the AIX Expansion Pack. This software can also be downloaded from the following sites.

OpenSSH:

<http://sourceforge.net/projects/openssh-aix>

OpenSSL:

<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=aixbp>

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

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

2. Download and install the prerequisite Open Source Software (OSS)

- Download the dep-aix.tar.gz tar file from <http://sourceforge.net/projects/xcat/> and copy it to a convenient location on your xCAT management node.
- Unwrap the tar file.

```
gunzip dep-aix.tar.gz  
tar -xvf dep-aix.tar.gz
```
- Read the README file.
- Run the **instoss** script (contained in the tar file) to install the OSS packages.

3. Verify that the following basic AIX services are configured and running:

- **inetd**
inetd includes services such as telnet, ftp, bootp, and others. Edit the /etc/inetd.conf file to turn on all services that are needed. Ftp and bootp are required for pSeries 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. A handy trouble-shooting guide listing the required services can be found at:

http://publib.boulder.ibm.com/infocenter/systems/index.jsp?topic=/com.ibm.aix.commadmn/doc/commadmndita/nfs_troublesh.htm

▪ **Name resolution**

Name resolution is required by xCAT. This cookbook assumes that you are using a simple `/etc/hosts` mechanism, but setting up DNS would also work. Make sure to verify and update the contents of both `/etc/hosts` (you may wish to use the xCAT `makehosts` command to do this after your nodes have been defined) and `/etc/resolv.conf` (for domain name and external name servers).

4. Download and install the xCAT software.

- Download the `core-aix.tar.gz` tar file from <http://sourceforge.net/projects/xcat/> and copy it to a convenient location on your xCAT management node.

- Unwrap the xCAT tar file.

```
gunzip core-aix.tar.gz  
tar -xvf core-aix.tar
```

- Run the `instxcat` script (contained in the tar file) to install the xCAT software. 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.)
- Execute the system profile file to set the xCAT paths. This file was updated during the xCAT post install processing.

```
./etc/profile
```

5. Verify the xCAT configuration.

- 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=no
xcatdport=3001
xcatiport=3002*

4.0 Deploying AIX nodes

4.1 Installing AIX nodes (using standard NIM rte method)

- add link to How-To

4.2 Booting AIX diskless nodes

- add link to How-To

4.3 Cloning AIX nodes using an AIX mksysb image

- add link to How-To - if I get to it!!!!

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/>