

XCAT 2 on AIX

Overview

3/26/2009

1.0 Release Description	1
1.1.1 xCAT for AIX Support Summary	2
1.1.2 Licensing	2
2.0 Overview of xCAT support for AIX	3
2.1 xCAT Commands	3
2.2 Database tables	3
2.3 Object definitions	3
2.4 Syslog setup	4
2.5 DNS setup	5
2.6 Remote shell setup	5
2.6.1 Setting up OpenSSH	5
2.7 NTP setup	6
2.8 Inventory	7
2.9 Parallel Commands	7
2.10 Customization script support	7
2.11 Operating system image management	8
2.12 Hardware discovery	8
2.13 LPAR management	9
2.14 Hardware control	9
2.15 Remote console support	9
2.16 Using xCAT service nodes in AIX clusters	10
2.17 xCAT Notification and Monitoring Plug-in Infrastructure	10
3.0 Installing xCAT and prerequisite Software	10
4.0 xCAT on AIX documentation	13
4.1 Installing AIX standalone nodes (using standard NIM rte method)	13
4.2 Booting AIX diskless nodes	13
4.3 Cloning AIX nodes (using an AIX mksysb image)	13
4.4 Using xCAT Service Nodes with AIX	13
5.0 References	13

1.0 Release Description

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

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

xCAT is a scalable distributed computing management and provisioning tool that provides a unified interface for hardware control, discovery, and OS diskful/diskless

deployment. Now xCAT 2 is open source on the [Source Forge Website](#) , so you can use it with confidence and participate in making it even better.

This document provides an overview of the xCAT 2.1 support for AIX clusters and instructions on how to install xCAT on AIX systems.

1.1 xCAT for AIX Support Summary

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

- Deploying AIX diskless and diskfull (standalone) nodes.
- Automation of AIX NIM setup.
- Support for AIX 5.3 and higher releases.
- Hierarchical support to allow large system to distribute the management of the cluster to service nodes. (xCAT 2.2 only)
- Operating system image management.
- Support for 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.
- Support for POWER 5, POWER 6, and JS blade (xCAT 2.2 only) 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.
- Software and firmware inventory.
- Allow continuous operation during cluster software updates using plug-in job scheduler (LoadLeveler, Moab).
- Automatic setup for syslog, remote shell, DNS, and ntp for both the xCAT management node, service nodes and the cluster nodes.
- xCAT cluster data stored in a relational database. (SQLite, MySQL,Postgresl, TBD)
- Commands to manipulate the xCAT database tables directly.
- Commands to manage xCAT data object definitions.
- Documentation and “**man**” pages.

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 all the supported xCAT commands. (Summaries and links to **man** pages.)

Note that some of the commands listed above are applicable only to xCAT for Linux support.

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.
- **boottarget** - Target profiles with their accompanying kernel parameters.
- **eventlog** - Stores the events occurred.

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 (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.

2.5 DNS setup

To setup the management node as the DNS name server you must set the “nameservers” and “forwarders” attributes in the xCAT “site” table.

For example, if the name of the management node is “mn20” and the site DNS servers are “9.14.8.1,9.14.8.2” then you would run the following command.

```
chdef -t site nameservers= mn20 forwarders= 9.14.8.1,9.14.8.2
```

Add the cluster node IP addresses and hostnames to the “/etc/hosts” file.

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

```
search <domain name>
nameserver mn20
```

Create xCAT networks definitions. Your network and mask value need to be defined, for **makedns** to use to setup the correct ip range for the management node to serve.

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

Start DNS:

```
startsrc -s named
```

2.6 Remote shell setup

The **xdsh**, **xdep**, or **xdshbak** commands require that you have either **ssh** or **rsh** configured. The **psh** command requires **ssh**. The xCAT on AIX support will automatically configure **rsh** by default. If you wish to have **ssh** configured, follow the instructions below.

2.6.1 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 postscripts 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.

2.7 NTP setup

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

Next set the “*ntpervers*” 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 ntpervers= 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.

```
chdef -t site ntpservers= mn20
```

To have xCAT automatically set up ntp on the cluster nodes you can add “*setupntp*” to the “*xcatdefaults*” entry in the *postscripts* table. The “*xcatdefaults*” entry is a list of xCAT customization scripts that are automatically run during node deployment. Add “*setupntp*” after the “*syslog*” entry. You can use the **chtab** or **tabedit** command to modify the *postscripts* table.

For example.

```
chtab node=xcatdefaults
postscripts.postscripts=syslog,aixremoteshell,setupntp
```

2.8 Inventory

The **sinv** command is designed to check the software and firmware configuration of nodes in a cluster or the diskless image. The command takes as input command line flags and one or more templates which will be compared against the output of a **xdsh** command run on the nodes. The nodes will then be grouped according to the template they matched and a report returned to the administrator. Whether root is required, depends on the required permission on the command to run.

See the **sinv** man page for parameters and examples.

2.9 Parallel Commands

The **xdsh** and **xdcp** command are provided to support parallel remote command execution and copy in the xCAT cluster. These commands are integrated with the xCAT database and daemon.

See the man pages for **xdsh** and **xdcp** for parameters and examples.

2.10 Customization script support

xCAT supports the automatic running of user-provided customization scripts on the nodes when they are deployed (installed or booted).

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.

2.11 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 (**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 definition 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.

2.12 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.
- **lssl** - 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 **lssl** 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.

2.13 LPAR management

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

You can:

- Create new diskless 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.)

2.14 Hardware control

The following xCAT 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.

2.15 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 consver:

Set the “cons” attribute of the node definitions to “hmc”.

```
chdef -t node -o aixnodes cons=hmc
```

Run the xCAT command.

```
makeconservercf
```

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

```
rcons node01
```

2.16 Using xCAT service nodes in AIX clusters

In an xCAT cluster the single point of control is the xCAT *management node*. However, in order to provide sufficient scaling and performance for large clusters, it may also be necessary to have additional servers to help handle the deployment and management of the cluster nodes. In an xCAT cluster these additional servers are referred to as *service nodes*.

This xCAT support requires xCAT version 2.2 or greater.

See the link below (in Section 4.4) for details on using service nodes in an AIX cluster.

2.17 xCAT Notification and Monitoring Plug-in Infrastructure

There are two monitoring infrastructures introduced in xCAT 2. The *xCAT Monitoring Plug-in Infrastructure* allows you to plug-in one or more third party monitoring software such as Ganglia, RMC, SNMP etc. to monitor the xCAT cluster. The *xCAT Notification Infrastructure* allows you to watch for the changes in xCAT database tables.

See “xCAT 2 Monitoring Howto” for details. (<http://xcat.svn.sourceforge.net/svnroot/xcat/xcat-core/trunk/xCAT-client/share/doc/xCAT2-Monitoring.pdf>)

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.)
- Apply the latest software updates and fixes if needed.
- 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>

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. Update the RPM repository

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.

```
updtvpkg
```

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

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

```
gunzip dep-aix-2.2.tar.gz  
tar -xvf dep-aix-2.2.tar
```
- Read the README file.
- Run the **instoss** script (contained in the tar file) to install the OSS packages.

4. 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. You can use 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).

5. Download and install the xCAT software.

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

```
gunzip core-aix-2.2.tar.gz  
tar -xvf core-aix-2.2.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*”). (**Note:** Make sure you don't have a .profile file that overwrites the “PATH” environment variables.)

6. 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 xCAT on AIX documentation

4.1 *Installing AIX standalone nodes (using standard NIM rte method)*

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

4.2 *Booting AIX diskless nodes*

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

4.3 *Cloning AIX nodes (using an AIX mksysb image)*

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

4.4 *Using xCAT Service Nodes with AIX*

(Requires xCAT version 2.2 or greater!)

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

4.5 *Updating AIX cluster nodes*

<http://xcat.svn.sourceforge.net/svnroot/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/>