XCAT 2 on AIX

Creating an xCAT Management Node

5/15/2009

1.0 Overiew

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

2.0 Pre-install configuration

2.1 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
```

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 files 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*".)

2.2 Clean up /etc/exports

If you are updating your management node with a new version of xCAT you should check the /etc/exports file to make sure that any old entries for /install/postscripts are removed.

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.   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.)

5.   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=no
    xcatdport=3001
    xcatiport=3002

4.0 Additional configuration of the management node

4.1 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 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".

4.2 Choose the shell to use in the cluster (optional)

By default the xCAT support will automatically set up **rsh** on all AIX cluster nodes. If you wish to use **ssh** you should modify the cluster site definition. To use **ssh** you would have to set the "useSSHonAIX=yes". 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".

For example:
    chdef -t site useSSHonAIX=yes

You will also have to make sure that the **openssl** an **openssh** software is installed on your nodes.

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

4.3 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.

4.3.1 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

4.3.2 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 all the 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.)

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

Start DNS:
startsrc -s named

4.4 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.

4.5 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*

4.6 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 attribute you want.

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'*

4.7 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*

4.8 Increase file size limit

Some of the /AIXNIM 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***

4.9 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*

4.10 Check system services

- **inted**

  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 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

5.0 xCAT on AIX documentation

5.1 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

5.2 Booting AIX diskless nodes

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

5.3 Cloning AIX nodes (using an AIX mksysb image)

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

5.4 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

5.5 Updating AIX cluster nodes

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

6.0 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/