System P Hardware Management

01/07/10, 10:55:42 AM

1. Introduction

This cookbook provides information and step-by-step instructions used with hardware discovery by the xCAT Management Node (MN), and setting up connections between HMC to IBM System P machines.

1.1 Terminology

The following terms will be used in this document:

- Frame (BPA) node: in xCAT 2.x, a node with nodetype set to *bpa* represents a high end System P server 24 inch frame.   For example, here is a frame node:

```
Object name: Server-9458-100-SN0SQIH42-1
    groups=bpa,all
    hcp=Server-9458-100-SN0SQIH42-1
    id=5
    mgt=bpa
    mtm=9458-100
    nodetype=bpa
    serial=0SQIH42
```

In above example, the attribute "nodetype" indicates this is a frame node, where attribute 'id' indicates the frame id number; The attribute 'mgt' indicates the current type of the hardware control point (hcp) is *bpa* , and attribute 'hcp' inidcates the hcp server node object. For this example the hcp is set to itself which is the default environment prior to be managed by any HMC. The attributes 'mtm' and 'serial' indicate the machine type model and its serial number of the server.

For lower end System P servers, there is no BPA device contained in a 19 inch frame, so there is no xCAT node object represented for 19 inch System P frame.

- CEC (FSP) node: in xCAT 2.x, has a server node with attribute nodetype set to *fsp* which represents a System P CEC. Here is an example of CEC node that exists in a high end System P server:

```
Object name: Server-9118-575-SN0SQIH42
    groups=all,fsp
    hcp=Server-9118-575-SN0SQIH42
    id=6
    mgt=fsp
```

```
mtm=9118-575
nodetype=fsp
parent=Server-9458-100-SN0SQIH42-1
serial=0SQIH42
```

In above example, the attribute 'nodetype' indicates this is a CEC node; and the'id' attribute indicates the cage number of this CEC in a 24 inch frame. (This will be set to blank for low end System P machine). The attribute 'parent' indicates the frame BPA server node that this CEC is controlled by. The 'mgt' attribute is initially set to *fsp* and 'hcp' attribute is set to itself as the CEC server node prior to be managed by the HMC.

2. System P hardware discovery

This chapter will introduce how the xCAT MN can discover HMCs, System P frames, and CECs working with xCAT "lsslp" command. The System P hardware will be discovered on the xCAT service network, and then added to xCAT database as node attributes.

2.1 Prerequisites

Before performing hardware discovery, users should confirm the following prerequisites:

2.1.1 site table:

Make sure the following attributes in "site" table are checked and are properly adjusted to match you xCAT cluster site environment:

- domain
- nameservers
- ntpservers

2.1.2 Network configuration

The xCAT MN needs to be properly connected to the xCAT service subnet which is used with all HMCs, System P frames and CECs being used in the xCAT cluster. This service subnet should be located on a private subnet to allow xCAT MN, DHCP server to communicate with HMCs, BPA (frame), and FSP (CECs) in your cluster. The following is the schematic diagram of the xCAT network topology:
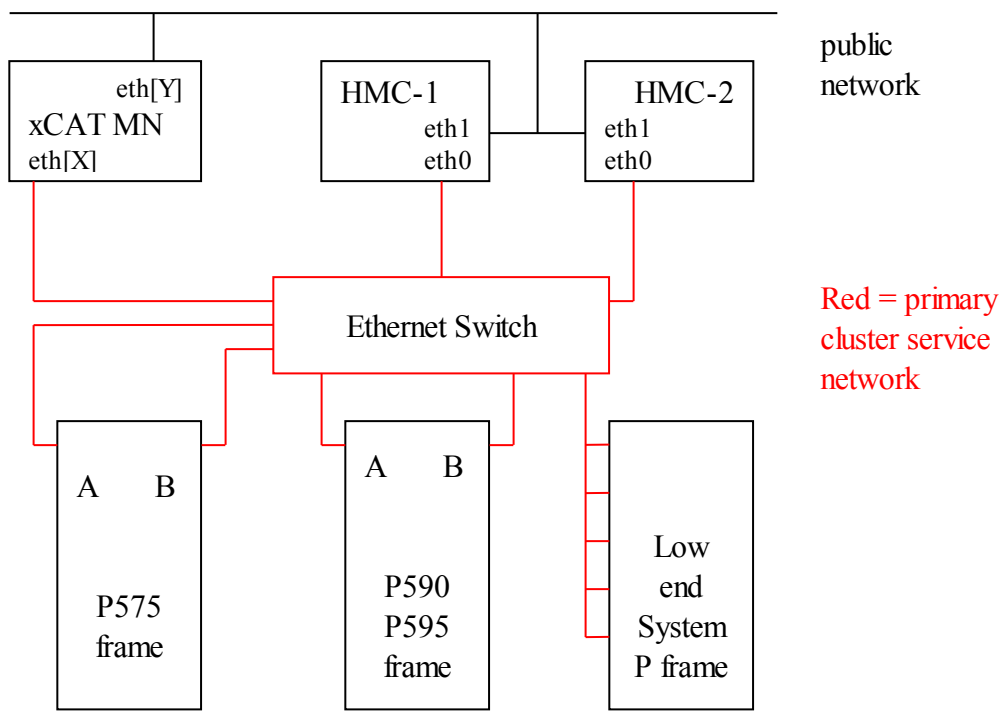
Chart 1.1 Cluster Service network topology

2.1.3 System P Hardware and HMC

- This function currently supports System P5/P6 hardware in xCAT 2.3 and xCAT 2.4.

- HMCs should be configured with correct static IP addresses in service subnet, so that they can communicate with xCAT MN. Because the DHCP service is not currently supported on the HMC, (DHCP service normally running on xCAT MN), the DHCP service on HMC should be turned off prior to performing this

discovery function. (By default, the DHCP service is disabled for all network interfaces on HMC.)

Reference the HMC website and documentation for more knowledge. The following are minimal steps required to Setup the HMC network for Static IP,and enable SLP and SSH ports working with HMC GUI.

a. Open the HMC GUI, Select **HMC Management**, then **Change Network Settings**.

b. Select **Customize Network Configuration**, and then **LAN Adapters** .

c. Select **Ethernet interface** configured on the service network.

d. Click on the **Details** button.

e. Select **Basic Settings**, Click on **Open**, and **Specify IP address.** Fill in **IP address**, **Netmask** for HMC static IP on the xCAT service network. Make sure that DHCP Server box is not selected and is blank.

f. Select on **Firewall Settings**, Click on **SLP, Secure Shell,** in the upper window.(You may also want to enable other HMC Firewall settings)

g. Click on the **Allow incoming** button for each required setting.

h. Make sure you Select OK at the bottom of the window to save your updates. Reboot the HMC, and then make sure Network changes are properly working.

- The DHCP service can be run from different server that is connected to the xCAT service subnet, instead of xCAT MN. In this case, users need to configure the DHCP service manually, and skip the step "Setup DHCP serice on MN" (section 2.2).

- The Frame and CEC should to be configured as dynamic IP so that the DHCP server can properly assign proper IP addresses in the xCAT service subnet. It admin wants to use static IP's with the BPA/FSP, it must be set in the proper dynamic address range specified with the DHCP server. For low end System P servers, only CEC/FSP have to be configured.)

2.1.4 Cleanup BPA/FSP IPs on Service Network and HMC

The xCAT administrator needs to make sure that the BPA/FSP IP addresses and server node names are properly understood working with the xCAT Database and the DHCP environment. There should be no implementation issues if this is a new System P cluster installation environment where the frames and CECs specified in the xCAT database or HMC for the first time.

For existing xCAT clusters previously setup with standard HMC DHCP environment where BPA/FSPs are already acknowledge by HMC and xCAT DB,

it is important that they use the same existing BPA/FSP network IP addresses and server node names. This includes setting up the DHCP server dynamic address ranges to match the current subnets used by the BPA/FSPs.

If the service network requires changes to the BPA/FSP IP addresses, the administrator should plan to cleanup the current BPA/FSP environment. This includes doing cleanup for both the HMC and the xCAT Database for any IP and server node name changes.

For the HMC, the administrator should plan to remove the existing Frames and Servers that will require new IP addresses, working in the new subnet. This will allow the xCAT "mkhwconn" command to reinitialize the frame and CECs used by the HMC.

For the xCAT MN, the administrator should review the xCAT DB using lsdef and tabdump commands to reference any existing HMC/frame/Server node objects that require updates. The xCAT chdef command can be used to modify server node attributes, or rmdef command if they want to remove the HMC/frame/Server node objects to get to a clean state. It is important that xCAT administrator also clean up Domain Name Service (DNS) and */etc/hosts* file make sure the HMC/frame/Server IP addresses and host names are matching the proper settings required for their xCAT cluster.

2.1.5 Hostname and IP address Planning

For xCAT hardware discovery feature, the prerequisite information is the hostnames and ip addresses allocation for all the FSPs and BPAs. The mapping between hostnames and ip addresses and the mapping between hostnames and physical hardware components has to be settled down before proceeding with the xCAT hardware discovery function.

Note: If it is a migration scenario such as migrating the CSM CRHS environment to xCAT and it is OK for you to use the default FSP/BPA host names, i.e., Server-*<Model>-<Type>-<SerialNumber>-<Side>*, then all the steps in 2.1.5 **"Planning"** section are not necessary at all, you can skip all the steps in sections **"Planning"**. If the FSPs and BPAs already have static ip addresses, then all the steps in confirgure DHCP section are not necessary. All the HMC connections have to be cleaned up and all the HMCs have to be changed back to standalone mode before proceeding with the xCAT hardware discovery.

Each FSP side and BPA side should have one ip address assigned, and each ip address should have a hostname defined, the FSPs and BPAs are indexed by hostnames.

The ip addresses for FSPs and BPAs are got from the DHCP server on xCAT management node, the DHCP service process in the hardware discovery is a little bit complex, there are two types of ip addresses allocation mechanism that can be

used in the hardware discovery process: random ip addresses or permanent ip addresses.

DHCP service can be configured to use dynamic ip range in the DHCP configuration file, or use the MAC addresses and ip addresses mapping in leases file. When using dynamic ip range, each DHCP client will be assigned a random ip address at first, after that, the previous ip address and the lease expire information will be used to determine if the DHCP client will get a new ip address when the DHCP client restarts. When using MAC addresses and ip addresses mapping in DHCP leases file, each DHCP client will get a specific ip address, we call it permanent ip address, if the DHCP client MAC address and ip address mapping is configured in the DHCP leases file, after that, the DHCP client will get the same ip address upon each restart.

Random ip addresses: When the FSPs and BPAs are brought up first time, since the MAC addresses for the FSPs and BPAs are not known yet, we can not specify the MAC address and ip address mapping in DHCP leases file, we can only specify a dynamic ip range in DHCP configuration file, so each FSP or BPA will get a random ip address. As mentioned above, the random ip address for each FSP or BPA may change in the future when the DHCP client on FSP or BPA restarts, however, if we specify a large enough dynamic ip range to aviod the ip addresses reuse, the random ip addresses can be similar as "permanent" ones. Please be aware that using the random DHCP ip addresses will increase the maintainance effort because you do not exactly know which hardware component has which ip address, at the same time, using random DHCP ip addresses opens an error windows that the FSPs/BPAs ip addresses may be changed during the FSPs/BPAs reboot, the FSPs/BPAs ip addresses change will result in HMC connection lost then you have to do some manual steps to recover. However, the random DHCP ip addresses solution should be able to work well for most of the scenarios.

Permanent ip addresses: the permanent ip addresses is clearer and easier to maintain, to achieve the permanent ip addresses goal, a method is to specify the MAC address and ip address mapping for each FSP and BPA in the DHCP leases file. Of course, getting the MAC address for each FSP and BPA is complex and time consuming unless the FSPs and BPAs have already got ip addresses from DHCP server, then we can simply ping the ip addresses and the mac addresses will be added to the arp cache. lsslp will collect the mac address and add the ip address mapping into DHCP leases file.

Either the solution "Random ip addresses" or solution "Permanent ip addresses" can be used, the administrator needs to select one solution as the ip addresses assignment mechanism.

The HMC host names should be resolvable through /etc/hosts, DNS or hosts table in xCAT DB before proceeding with the hardware discovery process.

2.2 xCAT DB setup and hardware discovery.

TO distinguish the randomly IP addresses and permanent IP addresses solution, we split the xCAT DB setup and hardware discovery steps into two scenarios, system administrators only need to look into the solution they select:

2.2.1 Randomly IP addresses

2.2.1.1 Setup DHCP server

1. **networks table**

All the FSPs and BPAs need to get their random IP addresses from DHCP server, the first thing to setup DHCP server is to create network table with mkdef command for the service network used by xCAT cluster.

Following is an example that provides random IP subnet *"192.168.200.0"*.

> *#netname,net,mask,mgtifname,gateway,dhcpserver,tftpserver,nameservers,
> ntpservers,logservers,dynamicrange,nodehostname,comments,disable*
>
> *"RandomIPRange","192.168.200.0","255.255.255.0","en0","192.168.200.205","1
> 92.168.200.205","192.168.200.205","192.168.200.205",,"192.168.200.205","192.
> 168.200.1-192.168.200.255",,,*

In the above example, network interface name en0 is used for random ip address, "192.168.200.1-192.168.200.255" field indicates the dynamic IP range used by the service network. The BPA/FSPs will receive dynamic IP addresses in this range from DHCP server 192.168.200.205 (which is the xCAT MN).

There is an xCAT command makenetworks that may be executed working with Linux xCAT MN that can populate xCAT networks table, but this command will not fill the field dynamicrange. The xCAT administrator needs to set the dynamicrange attribute in the network definitions working with xCAT chdef command according to their DHCP requirements .

On AIX xCAT MN, command makenetworks is not supported yet. All attributes fields in networks table need to be updated by the xCAT administrator working with xCAT chdef command.

2. **Stop bootp service and initialize DHCP service(AIX only)**

For Linux clusters, there is one xCAT command makedhcp to write dhcp service to inittab **(/etc/inittab)** automatically.

For AIX clusters, there is a bootp service used on the xCAT MN working with node installations. If the xCAT administrator wants to use DHCP service instead of bootp for xCAT cluster node installations, they will need to disable the bootp service and write dhcpsd into rc.tcpip to start dhcp service during system bootup. And also the NIM initialization will enable bootp automatically, stop bootp and re-enable dhcp

before creating your image and installing the nodes if you use dhcp for node installation.

Disable the bootp service (for AIX only, optional)

a) Stop bootp from rebooting by commenting bootps line in /etc/inetd.conf file:

> *#bootps dgram udp wait root /usr/sbin/bootpd bootpd /etc/bootptab*

b) Stop and restart the inetd subsystem:

> *refresh -s inetd*

c) Stop bootp deamon:

> *ps -ef | grep bootp*
>
> *kill the bootp process*

d) Start dhcpsd server from rebooting by uncomment dhcpsd lines in /etc/rc.tcpip

> *# Start up the DHCP Server*
>
> *start /usr/sbin/dhcpsd "$src_running"*

e) Stop and restart the tcpip group

> *stopsrc -g tcpip*
>
> *startsrc -g tcpip*

3.   Write static IP address for the network interface running for DHCP service.

For many reasons, xCAT Management Node could be rebooted, it is necessary to write a static IP address for the network interfaces on Management Node, so that DHCP server could provide service automatically after DHCP server reboot.

The following examples represent the DHCP network interface ent0 being used .

For Linux Clusters, write static IP address into network configuration file.

> For RHEL, the network configuration is in file like /etc/sysconfig/network-scripts/ifcfg-eth0:

> > *DEVICE=eth0*
> >
> > *BOOTPROTO=static*
> >
> > *HWADDR=00:14:5E:5F:20:90*

```
IPADDR=192.168.200.1
NETMASK=255.255.255.0
NETWORK=50.0.0.0
ONBOOT=yes
```

For SLES, the network configuration is in file like /etc/sysconfig/network/ifcfg-eth0:

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=00:14:5E:5F:20:90
IPADDR=192.168.200.1
NETMASK=255.255.255.0
STARTMODE=onboot
```

For AIX Clusters, issue the following commands to write static IP address for network interface:

```
mktcpip -a 192.168.200.1 -i en0 -m 255.255.255.0 -h xCAT_MN
```

4. Generate dhcp configuration file for service nodes and compute nodes.

There is xCAT command makedhcp that can be used to help to create the DHCP configuration file and establish the DHCP service daemon working on the xCAT MN.

Run "makedhcp -n" to add a dynamic IP pool.

Command makedhcp with flag -n will create the dhcp service configuration file based on attributes found in the xCAT site and networks tables. . In this configuration file, the dynamic address range IP pool is created based on the field dynamicrange in networks table.

Even if there are no definitions listed in networks table, the makedhcp command will try to generate generate DHCP service for all active subnets found on xCAT MN, even though there are no dynamic IP ranges.

The xCAT administrator should examine the DHCP configurations files on the xCAT MN /etc/dhcpd.conf (Linux) or /etc/dhcpsd.cnf (AIX) after execution of makedhcp. If you dont want the DHCP service to serve a subnet, you have to manually comment or remove the subnet definitions in /etc/dhcpd.conf (Linux) or /etc/dhcpsd.cnf (AIX) .

See makedhcp man page for details of this command.

2.2.1.2 Power on all the frames and CECs

Power on the frames/CECs manually, since the DHCP server is configured with dynamic ip range, then all the FSPs and BPAs will get floating IP addresses.

2.2.1.3 Update xCAT tables

Some information that mentioned in the "Planning" section should be added into xCAT DB, including the system specified node names, the mapping between the ip addresses and hostsnames, the mapping between the hostnames and the physical hardware components.

1. Update the nodelist table

The **"nodelist"** table stores all the hostnames and groups information for all the nodes. Before proceeding with the hardware discovery process, the "nodelist" table needs to be updated to include all the FSPs and BPAs hostnames information.

- *nodelist table*

  #node,groups,status,appstatus,primarysn,comments,disable

  frame1a,"all,bpa,bpa-a"

  frame1b,"all,bpa,bpa-b"

  ... ...

  frame16a,"all,bpa,bpa-a"

  frame16b,"all,bpa,bpa-b"

  f1c1,"all,fsp"

  f1c2,"all,fsp"

  ... ...

  f16c16,"all,fsp"

In the nodelist table, each node may have different status and appstatus in the future, so we could not use "groups" in nodelist table, each node has to be listed separately in the nodelist table. It is time consuming to input all the nodes into nodelist table, the mkdef command can be used to add all the nodes nodelist table:

mkdef frame1a-frame16a groups="all,bpa,bpa-a"

mkdef frame1b-frame16b groups="all,bpa,bpa-b"

2. Update the mapping between the hostnames and physical hardware components

1). Use vpd mode

For the high end servers such as POWER 595, only the BPAs MTMS information is necessary, for the low end servers such as POWER 520, the FSPs MTMS information is needed. System administrators have two options to get the MTMS information for all the BPAs or FSPs, one is to copy the BPAs or FSPs MTMS information from the front rear of the BPA or CEC manually; or run lsslp to get the basic hardware information from the lsslp output.

Run *lsslp -s BPA -z -i 10.0.0.1 > /stanza/file/path*

or

Run *lsslp -s FSP -z -i 10.0.0.1 > /stanza/file/path*

The vpd method uses the MTMS and parent information to determine which hardware component that the hostname is pointing to. For high end servers environment, the "vpd" table needs to be updated to include the BPA MTMS information and the FSP side information, if the systems only have one FSP for each CEC, then the FSP side information is not necessary in vpd table. For the low end servers environment, the "vpd" table needs to be updated to include the FSP MTMS information. All the MTMS information for the BPAs or FSPs can be got from the stanza file */stanza/file/path*.

- vpd table:

  #node,serial,mtm,side,asset,comments,disable

  "frame1a","99200G1","9A00-100","A",,,

  "frame1b","99200G1","9A00-100","B",,,

  ... ...

  "frame16a","99410D1","9A00-100","A",,,

  "frame16b","99410D1","9A00-100","B",,,

Note: The frames' MTMS information can be got from the stanza file.

For low end servers, the FSPs MTMS information can be got from the stanza file.

For high end CECs, the "ppc" table also needs to be updated to include the cage id information and the BPA information for each FSP. For low end CECs, the "ppc" table does not need to be updated.

- ppc table:

  #node,hcp,id,pprofile,parent,supernode,comments,disable

  "f1c1",,"1",,"frame1a",,,

  "f1c2",,"2",,"frame1a",,,

  ... ...

  "f16c1",,"1",,"frame16a",,,

```
"f16c2",,"2",,"frame16a",,,
```

2). Use the switch port mode

The switch port method uses the switch connection information to determine which hardware component that the hostname is pointing to. The "switch" table needs to be updated to indicate which switch port the FSP/BPA connects to. For high end CECs, only BPA/frame hostnames need to be added into switch table. For low end CECs, only CEC/fsp hostnames need to be added in. For example:

- switch table

```
#node,switch, port,vlan,interface,comments,disable
"frame1a","10.0.0.254","1",,,,
"frame1b","10,0.0.254","2",,,,
      ... ...
"frame16a","10.0.0.254","31",,,,
"frame16b","10.0.0.254","32",,,,
```

For high end CECs, the "ppc" table also needs to be updated to include the cage id information for each FSP. For low end CECs, the "ppc" table does not need to be updated.

- ppc table

```
#node,hcp,id,pprofile,parent,supernode,comments,disable
"f1c1",,"1",,"frame1a",,,
"f1c2",,"2",,"frame1a",,,
        ... ...
"f16c1",,"1",,"frame16a",,,
"f16c2",,"2",,"frame16a",,,
```

Note that if there is hub during the connection between HMC and BPA/CEC, the BPA/CEC MTMS information should also been added into vpd table, see section "Use vpd mode" for detailed description.

2.2.1.4 Discover HMCs/frame/CECs, and define them in xCAT DB.

 XCAT command lsslp will help to discovery the HMC/frame/CECs, write the discovered information into xCAT DB or output in different format, including RAW, XML and stanza format.

1. Use stanza file

1). Issue lsslp to get the HMC information and write into stanza file
lsslp -s HMC -z -i 10.0.0.1 > /hmc/stanza/file

2). Review the HMC stanza file and make modifications if necessary.
3). Write the HMC information into xCAT DB with xCAT command mkdef.
cat /hmc/stanza/file | mkdef -z

4). Issue lsslp to get BPA information in switchport or vpd mode and write into stanza file.
lsslp -s BPA -z -i 10.0.0.1 -M {switchport|vpd} --updatehosts > /bpa/stanza/file

5). Review the BPA stanza file and make modifications if necessary.
6). Write the BPA information into xCAT DB with xCAT command mkdef
cat /bpa/stanza/file | mkdef -z

7). Issue lsslp to get FSP information in switchport or vpd mode and write into stanza file.
lsslp -s FSP -z -i 10.0.0.1 -M {switchport|vpd} --updatehosts > /bpa/stanza/file

8) Review the BPA stanza file and make modifications if necessary.
9) Write the FSP information into xCAT DB with xCAT command mkdef
cat /bpa/stanza/file | mkdef -z

2. Update xCAT database directly

1).Issue lsslp to get HMC information and write into xCAT DB directly.
lsslp -s HMC -w -i 10.0.0.1

2).Issue lsslp to get BPA information in switchport or vpd mode and write into xCAT DB directly.
lsslp -s BPA -w -i 10.0.0.1 -M {switchport|vpd} --updatehosts

3).Issue lsslp to get FSP information in switchport or vpd mode and write into xCAT DB directly.
lsslp -s FSP -w -i 10.0.0.1 -M {switchport|vpd} --updatehosts

Note: The lsslp -n option could help to discovery the newly added HMC/BPA/FSPs, the existing hardwares defined in xCAT DB will not be output by lsslp command with -n option. See man page of lsslp for details.

2.2.2 Permanent ip addresses

2.2.2.1 Setup DHCP server

1. networks table

The xCAT DB requires network tables to be created using the mkdef command for each subnet that is being used by the xCAT cluster.

All the FSPs and BPAs ip addresses are got from DHCP server on xCAT management node or standalone DHCP server and this random IP addresses tend to be changed dynamically, the random FSP/BPA ip addresses are hard to maintain and error prone, so permanent ip addresses is a more clear and easier solution. To achieve the permanent ip addresses goal, a method is to specify the MAC address and ip address mapping for each FSP and BPA in the DHCP leases file.

To avoid conflicting between the "random ip addresses" and "permanent ip addresses", two separate subnets are recommended. The following is an example that provides random IP subnet *"192.168.200.0"* and permanent IP subnet *"10.0.0.0"* in the xCAT networks table:

> *#netname,net,mask,mgtifname,gateway,dhcpserver,tftpserver,nameservers,*
> *ntpservers,logservers,dynamicrange,nodehostname,comments,disable*
>
> *"RandomIPRange","192.168.200.0","255.255.255.0","en0","192.168.200.205","1*
> *92.168.200.205","192.168.200.205","192.168.200.205",,"192.168.200.205","192.*
> *168.200.1-192.168.200.255",,,*
>
> *"PermanentIPAddress","10.0.0.0","255.255.255.0","en1","10.0.0.1","10.0.0.1","*
> *10.0.0.1","10.0.0.1",,"10.0.0.1",,,,*

In the above example, network interface name en0 is used for random ip address, "192.168.200.1-192.168.200.255" field indicates the dynamic IP range used by the service network. The BPA/FSPs will receive dynamic IP addresses in this range from DHCP server 192.168.200.205 (which is the xCAT MN).

Network interface name en1 is used for permanent ip address, "10.0.0.1" is the DHCP server, which is also the xCAT MN. There is no dynamicrange field put in this subnet since the BPA/FSPs should not receive dynamic IP addresses in this range.

There is an xCAT command makenetworks that may be executed working with Linux xCAT MN that can populate xCAT networks table, but this command will not fill the field dynamicrange. The xCAT administrator needs to set the dynamicrange attribute in the network definitions working with xCAT chdef command according to their DHCP requirements .

On AIX xCAT MN, command makenetworks is not supported yet. All attributes fields in networks table need to be updated by the xCAT administrator working with xCAT chdef command.

2.   **Stop bootp service and initialize DHCP service(AIX only)**

For Linux clusters, there is one xCAT command makedhcp to write dhcp service to inittab **(/etc/inittab)** automatically.

For AIX clusters, there is a bootp service used on the xCAT MN working with node installations. If the xCAT administrator wants to use DHCP service instead of bootp

for xCAT cluster node installations, they will need to disable the bootp service and write dhcpsd into rc.tcpip to start dhcp service during system bootup.

Disable the bootp service (for AIX only, optional)

1). Stop bootp from rebootting by commenting bootps line in /etc/inetd.conf file:

#bootps dgram udp wait root /usr/sbin/bootpd bootpd /etc/bootptab

2). Stop and restart the inetd subsystem:

refresh -s inetd

3). Stop bootp deamon:

ps -ef | grep bootp

kill the bootp process

4). Start dhcpsd server from rebootting by uncomment dhcpsd lines in /etc/rc.tcpip

 # Start up the DHCP Server

start /usr/sbin/dhcpsd "$src_running"

5). Stop and restart the tcpip group

stopsrc -g tcpip

startsrc -g tcpip

3.   Write static IP address of the network interfaces running for DHCP service.

For many reasons, xCAT Management Node could be rebooted, it is necessary to write a static IP address for the network interfaces on Management Node, so that DHCP server could provide service automatically after Management Node reboot.

The following examples represent the DHCP network interface ent0 being used .

For Linux Clusters, write static IP address into network configuration file.

For RHEL, the network configuration is in file like /etc/sysconfig/network-scripts/ifcfg-eth0:

DEVICE=eth0

BOOTPROTO=static

HWADDR=00:14:5E:5F:20:90

IPADDR=192.168.200.1

```
NETMASK=255.255.255.0
NETWORK=50.0.0.0
ONBOOT=yes
```

For SLES, the network configuration is in file like /etc/sysconfig/network/ifcfg-eth0:

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=00:14:5E:5F:20:90
IPADDR=192.168.200.1
NETMASK=255.255.255.0
STARTMODE=onboot
```

For AIX Clusters, issue the following commands to write static IP address for network interface:

```
mktcpip -a 192.168.200.1 -i en0 -m 255.255.255.0 -h xCAT_MN
```

4. Generate dhcp configuration file for service nodes and compute nodes.

There is xCAT command makedhcp that can be used to help to create the DHCP configuration file and establish the DHCP service daemon working on the xCAT MN.

Run "makedhcp -n" to add a dynamic IP pool.

Command makedhcp with flag -n will create the dhcp service configuration file based on attributes found in the xCAT site and networks tables. . In this configuration file, the dynamic address range IP pool is created based on the field dynamicrange in networks table.

Even if there are no definitions listed in networks table, the makedhcp command will try to generate generate DHCP service for all active subnets found on xCAT MN, even though there are no dynamic IP ranges.

The xCAT administrator should examine the DHCP configurations files on the xCAT MN /etc/dhcpd.conf (Linux) or /etc/dhcpsd.cnf (AIX) after execution of makedhcp. If you dont want the DHCP service to serve a subnet, you have to manually comment or remove the subnet definitions in /etc/dhcpd.conf (Linux) or /etc/dhcpsd.cnf (AIX) .

See makedhcp man page for details of this command.

2.2.2.2 Power on all the frames and CECs

Power on the frames/CECs manually, since the DHCP server is configured with dynamic ip range, then all the FSPs and BPAs will get floating IP addresses.

2.2.2.3 Update xCAT tables

Some information that mentioned in the "Planning" section should be added into xCAT DB, including the system specified node names, the mapping between the ip addresses and hostsnames, the mapping between the hostsnames and the physical hardware components.

1. Update the mapping between hostnames and ip addresses

1). Update nodelist table

The **"nodelist"** table stores all the hostnames and groups information for all the nodes. Before proceeding with the hardware discovery process, the "nodelist" table needs to be updated to include all the FSPs and BPAs hostnames information.

- nodelist table

  #node,groups,status,appstatus,primarysn,comments,disable

  frame1a,"all,bpa,bpa-a"

  frame1b,"all,bpa,bpa-b"

  ... ...

  frame16a,"all,bpa,bpa-a"

  frame16b,"all,bpa,bpa-b"

  f1c1,"all,fsp"

  f1c2,"all,fsp"

  ... ...

  f16c16,"all,fsp"

In the nodelist table, each node may have different status and appstatus in the future, so we could not use "groups" in nodelist table, each node has to be listed separately in the nodelist table. It is time consuming to input all the nodes into nodelist table, the mkdef command can be used to add all the nodes nodelist table:

  mkdef frame1a-frame16a groups="all,bpa,bpa-a"

  mkdef frame1b-frame16b groups="all,bpa,bpa-b"

2). Update hosts table

The "hosts" table stores the hostnames and permanent ip addresses for all the nodes. For hardware discovery scenario, the hostnames and ip addresses information for all the FSPs and BPAs should be in "hosts" table before proceeding with the hardware discovery process.

- hosts table

```
#node,ip,hostnames,comments,disable
"f1c1","10.0.0.2",,,
"f1c2","10.0.0.3",,,

   ... ...
"frame1a","10.0.0.127",,,
"frame1b","10.0.0.128",,,
```

Note: If the CECs have two FSPs installed such as the POWER 595 systems, both the two FSPs should be added into the xCAT database. Here is an example:

```
#node,ip,hostnames,comments,disable
"f1c1-a",  "10.0.0.2"
"f1c1-b",  "10.0.0.3"

   ... ...
"f16c1-a",  "10.0.0.127"
"f16c1-b",  "10.0.0.128"
```

2. Update the mapping between the hostnames and physical hardware components

1). Use vpd mode

For the high end servers such as POWER 595, only the BPAs MTMS information is necessary, for the low end servers such as POWER 520, the FSPs MTMS information is needed. System administrators have two options to get the MTMS information for all the BPAs or FSPs, one is to copy the BPAs or FSPs MTMS information from the front rear of the BPA or CEC manually; or run lsslp to get the basic hardware information from the lsslp output.

Since all the FSPs and BPAs have got ip addresses though the DHCP server are random ones, the lsslp command can be run to get the basic hardware information, but not all the necessary information is available such as the hostname and ip addresses mapping.

> *Run lsslp -s BPA -z -i 10.0.0.1 > /stanza/file/path*

or

> *Run lsslp -s FSP -z -i 10.0.0.1 > /stanza/file/path*

The vpd method uses the MTMS and parent information to determine which hardware component that the hostname is pointing to. For high end servers environment, the "vpd" table needs to be updated to include the BPA MTMS information and the FSP side information, if the systems only have one FSP for each CEC, then the FSP side information is not necessary in vpd table. For the low end servers environment, the "vpd" table needs to be updated to include the FSP MTMS information. All the MTMS information for the BPAs or FSPs can be got from the stanza file */stanza/file/path*.

- vpd table:

> *#node,serial,mtm,side,asset,comments,disable*
>
> *"frame1a","99200G1","9A00-100","A",,,*
>
> *"frame1b","99200G1","9A00-100","B",,,*
>
> *... ...*
>
> *"frame16a","99410D1","9A00-100","A",,,*
>
> *"frame16b","99410D1","9A00-100","B",,,*

Note: The frames MTMS information can be got from the stanza file. For low end servers, the FSPs MTMS information can be got from the stanza file.

For high end CECs, the "ppc" table also needs to be updated to include the cage id information and the BPA information for each FSP. For low end CECs, the "ppc" table does not need to be updated.

- ppc table:

> *#node,hcp,id,pprofile,parent,supernode,comments,disable*
>
> *"f1c1",,"1",,"frame1a",,,*
>
> *"f1c2",,"2",,"frame1a",,,*
>
> *... ...*
>
> *"f16c1",,"1",,"frame16a",,,*
>
> *"f16c2",,"2",,"frame16a",,,*

2). Use the switch port mode

The switch port method uses the switch connection information to determine which hardware component that the hostname is pointing to. The "switch" table needs to be updated to indicate which switch port the FSP/BPA connects to. For high end CECs, only BPA/frame hostnames need to be added into switch table. For low end CECs, only CEC/fsp hostnames need to be added in. For example:

- switch table

  #node,switch, port,vlan,interface,comments,disable

  "frame1a","10.0.0.254","1",,,,

  "frame1b","10,0.0.254","2",,,,

      ... ...

  "frame16a","10.0.0.254","31",,,,

  "frame16b","10.0.0.254","32",,,,

For high end CECs, the "ppc" table also needs to be updated to include the cage id information for each FSP. For low end CECs, the "ppc" table does not need to be updated.

- ppc table

  #node,hcp,id,pprofile,parent,supernode,comments,disable

  "f1c1",,"1",,"frame1a",,,

  "f1c2",,"2",,"frame1a",,,

       ... ...

  "f16c1",,"1",,"frame16a",,,

  "f16c2",,"2",,"frame16a",,,

Note that if there is hub during the connection between HMC and BPA/CEC, the BPA/CEC MTMS information should also been added into vpd table, see section 2.7.1 "Use vpd mode" for detailed description.

2.2.2.4 Discover HMCs/frame/CECs, and define them in xCAT DB.

XCAT command lsslp will help to discovery the HMC/frame/CECs, write the discovered information into xCAT DB or output in different format, including RAW, XML and stanza format.

1. **Use stanza file**

1). Discovery HMC firstly and write the information into stanza file.

```
lsslp -s HMC -z -i 10.0.0.1 > /hmc/stanza/file
```

2). Review the HMC stanza file and make modifications if necessary.

3). Write the HMC information into xCAT DB with xCAT command mkdef.

```
cat /hmc/stanza/file | mkdef -z
```

4). Discover BPA in switchport or vpd Mode and write into stanza file.

```
lsslp -s BPA -z -i 10.0.0.1 -M {switchport|vpd} > /bpa/stanza/file
```

5). Review the BPA stanza file and make modifications if necessary.

6). Write the BPA information into xCAT DB with xCAT command mkdef.

```
cat /bpa/stanza/file | mkdef -z
```

7). Discovery FSP in switchport or vpd Mode and write into stanza file

```
lsslp -s FSP -z -i 10.0.0.1 -M {switchport|vpd} > /bpa/stanza/file
```

8). Review the BPA stanza file and make modifications if necessary.

9). Write the FSP information into xCAT DB with xCAT command mkdef.

```
cat /bpa/stanza/file | mkdef -z
```

10). Issue xCAT command makedhcp to write permanent IP address into DHCP lease file.

```
makedhcp -a
```

11). Issue xCAT command lsslp with option resetnet to restart BPA's network to get permanent IP address from DHCP server.

```
lsslp -s BPA –resetnet
```

12). Issue xCAT command lsslp with option resetnet to restart FSP's network to get permanent IP address from DHCP server.

```
lsslp -s FSP –resetnet
```

2. Update xCAT DB directly

1). Dicovery HMC and write the information into xCAT DB directly with -w option

```
lsslp -s HMC -w -i 10.0.0.1
```

2). Discovery BPA and write the information into xCAT DB directly with -w option.

```
lsslp -s BPA -w -i 10.0.0.1 -M {switchport|vpd} --makedhcp --resetnet
```

3). Discovery FSP and write the information into xCAT DB directly with -w option.

```
lsslp -s FSP -z -i 10.0.0.1 -M {switchport|vpd} --makedhcp --resetnet
```

Note: The lsslp -n option could help to discovery the newly added HMC/BPA/FSPs, the existing hardwares defined in xCAT DB will not be output by lsslp command with -n option. See man page of lsslp for details.

2.3 Limitations:

- In a cluster that contains a large number of P5 575 machines, the "lsslp" command may not be able to discover all machines. You can reduce this scaling issue with lsslp by using the "-t" (retry times) and "-c" (timeout value) flags . For an example:

  lsslp -s FSP -i 192.168.200.246 -t 5 -c 3000,3000,3000,3000,3000

  See "lsslp" man page for the details.

- For HMC with V7R350 release, we had experienced some HMC discovery issues "*lsslp -m*" in different layer2/layer3 ethernet switch environments. In this case, the xCAT admin have to manually create the HMC server node object using xCAT command "*mkdef*".

- If you run xCAT command lsslp with flag "-w" to auto discover BPA/FSP and create BPA/FSP nodes in xCAT DB, there are some types of BPA/FSP that cannot respond the user-defined BPA/FSP system names to xCAT. This causes the node name created by lsslp to not be consistent to the system name that is known by HMC. This limitation will not block most functions of xCAT. If system admins want to sync the user-defined system names used by the HMC to xCAT DB, please run rscan with -u option to update the FSP/BPA node names in the xCAT database. The rscan -u command should only be executed after the running of the mkhwconn command (described in section 3).

3. Setup connections from HMC to frame/CEC

The xCAT administrator can setup the xCAT cluster to connect the BPA or FSP to selected HMCs that are located in their xCAT service network. They can also specify the proper security environment used with passwords to HMC, BPA, and FSP.

3.1 Set proper passwords for BPA/FSP/HMC.

The passwords used with BPA/FSP userid 'HMC' needs to be set correctly in xCAT table "ppcdirect" or table "passwd" if the cluster is not going to use the default password. Here is an example of table "ppcdirect",

> *#hcp,username,password,comments,disable*
>
> *"Server-9125-F2A-0262672","HMC","abc123",,*
>
> *"my_frame_grp1","HMC","abcdefg",,*

The default passwords of BPA/FSP for userid 'HMC' can also be set in table "passwd" which is used with xCAT cluster, for an example:

#key,username,password,comments,disable

"fsp","HMC","ibm4you",,

The passwords used with the HMC nodes working with userid "hscroot" is located in the xCAT table "ppchcp" . For an example

#hcp,username,password,comments,disable

"c76v1hmc02","hscroot","abc123",,

3.2 Run mkhwconn to assign frames/CECs to correct HMCs

The "mkhwconn" command allows the xCAT administrator to properly setup the FSP/BPA connection between the HMC and frames/CECs. This command will make the proper connections on the target HMC if the frame/CEC is not already connected. It also makes the proper updates to the xCAT DB working with the "ppc" and "nodehm" tables and with the target FSP/BPA node objects.

For an mkhwconn example, run

mkhwconn FSP1 –p HMC1 -P <HMC passwd>

will result with FSP node *FSP1* to be connected by HMC node *HMC1*.

See mkhwconn man page for details of this command

Note:

- Οnly one HMC can be specified and active in xCAT DB at a time.

- For those high-end System P servers (IH and H servers), the CEC/FSP nodes in the selected frame/BPA are controlled by the BPA node. In this case, the FSP nodes are not allowed to be assigned directly to an HMC node. The BPA nodes should be assigned to the HMC node, and the "mkhwconn" will assign all the FSP nodes in the selected frame to the HMC automatically.

If you need to assign FSP/BPA server nodes to multiple HMC nodes at a given time, you need to first define the second HMC connection into the xCAT database where nodehm and ppc tables, and FSP/BPA node object are updated with the chdef command. You then can run "mkhwconn" with "–t" flag to setup multiple BPA/FSP node to create a new connection to second HMC.

The following is an example of activating a second HMC:

Run chdef to change the "mgt" (node management type) to "hmc", and change "hcp" to the corresponding HMC node.

> *chdef frame_group1,fsp_group2 mgt=hmc hcp=hmc1.*

Here the *frame_group1* includes the selected BPA nodes, and *fsp_group2* includes FSP nodes that do not belong to the BPA nodes in frame_group1.

Run mkhwconn to have a second hmc1 be the second and main HMC connections for *frame_group1* and f*sp_group2.*

> *mkhwconn frame_group1,fsp_group2 –t*

Note: There can be only one active HMC defined in the xCAT DB working with a BPA/FSP server node at this time. The xCAT administrator can toggle between the two HMCs by using the chdef command working with the "mgt" and "hcp" attributes for the target BPA/FSP node.

3.3 List frames/CECs from HMC

There is the "lshwconn" command that will provide the current frame/CEC connection data that is specified on a HMC, BPA, or FSP. This information currently provides the FSP/BPA server node, the FSP/BPA IP address, and the connection status of the BPA/FSP used for the target HMC node.

Run *lshwconn <HMC node>* to locate all BPA/FSP servers on HMC.

See "lshwconn" man page for the details.

3.4 Move frames/CECs to other HMCs (optional)

There are times when you will want to remove frame/CECs from one HMC and then connect them to a different HMC. You can use the "rmhwconn" command to remove the BPA/FSP connection from the current HMC. You can then execute the "mkhwconn" command to generate a new connection to a second HMC.

For example, if the attribute "hcp" attribute is set to HMC node *hmc1*, working with BPA server node *frame1* running   "rmhwconn frame1"

This will remove the frame *frame1*, as well as any CECs located in the frame1 from current HMC *hmc1*. The rmhwconn will also reset the xCAT database to be in a state prior to initial HMC onnection (mkhwconn) . The BPA/FSP server node attributes "mgt" are set back to fsp or bpa, and "hcp" now points back to parent FSP/BPA server node instead of the HMC node.

Run mkhwconn to assign frames/CECs to another HMC, as described in 3.2.