

ISBN 978-89-6211-038-8 93560

슈퍼컴퓨터 4호기(대용량컴퓨팅시스템) 클러스터 관리  
시스템 구축 가이드

최윤근, 장지훈

# 목 차

1. CSM 이란	1
1.1 CSM 개요	1
1.2 CRHS	2
1.3 CSM hardware control	4
1.3.1 Security	5
2. CSM 설치 준비	6
2.1 cluster hardware 준비	6
2.1.1 CSM Management server 준비	6
2.1.2 CSM Managed node 준비	6
2.2 Cluster Software 준비	7
3. CSM 설치	10
3.1 Node hostname 정의	10
3.2 PATH 및 MANPATH 설정	11
3.3 설치관련 파일 시스템 생성	11
3.4 클러스터 소프트웨어 설치	11
3.5 Remote shell 속성 설정	12
3.6 License 인증	13
3.7 Hardware control point userID and password	13
3.8 CRHS setup	14
3.8.1 setup DHCP server on Management server	15
3.8.2 setup HMC cluster peer Domain	20
3.9 CSM 설치 확인	28
3.10 Kerberos setup	29
3.11 Installation server setup	30
3.11.1 노드 정의	31
3.11.2 Device 정의	33
3.12 NIM installation	33

3.12.1	각 노드의 정의 확인	34
3.12.2	Remote power control check	35
3.12.3	Network Adapter 정보 수집	35
3.12.4	기본 NIM setup	37
3.12.5	NIM mkysyb resource 생성	37
3.12.6	NIM machine 정의	38
3.12.8	AIX 5L 인증방식 선택	39
3.12.9	클러스터 내에 노드 추가 준비	39
3.12.9	노드 bos_inst operation	40
3.12.10	노드 설치	40
3.12.11	설치 후 점검	41
4.	HPS setup	43
4.1	고려 사항	43
4.1	용어 설명	43
4.3	KISTI HPS 구성	46
4.4	HPS network setup	49
4.4.1	HPS service login ID 생성	49
4.4.2	HMC에서 CMS management server 추가	50
4.4.3	HPSNM 시범 구동	50
4.4.4	HPSNM 정성 구동	54
4.4.5	HPSNM 명령어	54

## 1. CSM 이란?

### 1.1 CSM 개요

Cluster Systems Management (CSM)은 AIX 5L 혹은 Linux 시스템으로 이루어진 클러스터를 구성하고 관리하기 위한 소프트웨어로써 single point-of-control을 제공하여 손쉽게 클러스터의 자원을 통제 가능하게 한다. CSM 클러스터의 운영, 관리, 모니터링을 위해 반드시 CSM Management Server를 구성해야 하며, CSM Management Server를 통해 single point-of-control이 가능하다. CSM 클러스터 내에서 CSM Management Server의 통제를 받는 각 node들을 Managed Node라 부르며, Managed Node는 AIX 5L 혹은 Linux 시스템이 포함될 수 있다.

CSM Management Server가 수행하는 일을 대략적으로 살펴보면 아래와 같다:

- ▶ Managed Node 상에 OS 및 CSM, RSCT등 기타 소프트웨어 설치
- ▶ Distributed Command 수행 (Kerberos)
- ▶ 파일 동기화 (Configuration File Manager)
- ▶ Managed Node 모니터링 및 관리 (Resource Monitoring and Control)
- ▶ 클러스터 내부의 하드웨어 제어 (Cluster-Ready Hardware Server)

- ▶ Diagnostic tool 수행 (예:diagnostic probes)
- ▶ 추가 네트워크 어댑터 구성 (예:HPS)

CSM 클러스터를 구성하기 위해서 필요한 Network은 아래와 같다:

- ▶ Management VLAN : Management Server와 HMC를 연결
- ▶ Cluster VLAN : Management Server와 Managed Node를 연결

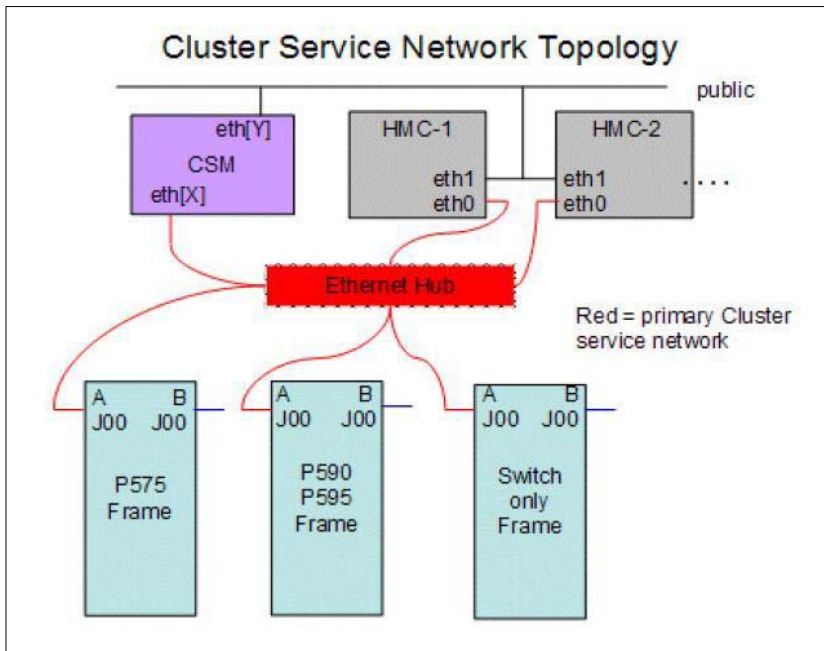
## 1.2 Cluster-Ready Hardware Server (CRHS)

CSM 클러스터에서 High Performance Switch (HPS)를 사용하기 위해서는 반드시 CRHS가 구성이 되어야 한다. CRHS는 Management VLAN을 이용하여 H/W control을 가능하게 한다.

CRHS가 제공하는 기능을 살펴보면 아래와 같다:

- ▶ Hardware 탐색 (hardware server discovery agent)
- ▶ 클러스터 Hardware 정보를 관리 (shared repository)
- ▶ 추가된 hardware server daemon (hdwr\_svr)
- ▶ Managed Node의 설치 및 FSP/BPA password setup (systemid)

CRHS service network은 management VLAN을 사용하며 아래와 같이 구성된다:



CRHS service network은 HMC를 사용하는 IBM system p 장비에 대해서는 반드시 구성이 되어져야하며, FSP, BPA, HMC 그리고 CSM management server가 연결되어야 한다. HMC는 반드시 eth0 인터페이스로 연결이 되어야 하지만, CSM은 제한이 없다.

먼저, hardware server discovery agent를 사용하면 CRHS service network에 연결되어 있는 모든 장비를 탐색하게 되며, 이 정보는 shared repository에 저장된다. 그 정보를 이용하여 HMC와 server 간의 상관관계를 만들어주면, CRHS 구성할 때 자동으로 management server와 HMC에 추가되는 "hdwr\_svr" daemon이 CRHS service network내의 모든 system p server 및 Frame BPA

와의 통신을 담당하게 된다.

### 1.3 CSM Hardware Control

3호기에서 사용중인 PSSP와 마찬가지로, CSM도 Management Server 상에서 Managed Node에 대한 하드웨어 원격 제어가 가능하다. PSSP와 마찬가지로, openCIMON을 Management Server에 설치하여야 원격 제어가 가능해 진다. 특히, Management Server에 AIX 5L이 탑재된 경우, 지원되는 모든 System p/x 노드에 대한 하드웨어 제어가 가능하다.

몇 가지 하드웨어 제어 명령어를 살펴보면 아래와 같다:

- ▶ addpeer : HMC를 사용하여 cluster peer domain을 구성한다.
- ▶ chrhws : shared repository내의 하드웨어 정보나 링크를 변경한다.
- ▶ getadapters : LAN adapter 정보를 수집한다.
- ▶ hwsda : 특정 네트워크 인터페이스에 대한 하드웨어를 검색한다.
- ▶ lshwdev : CSM DB 상에 정의된 device를 보여준다.
- ▶ lshwinfo : 하나 또는 다수의 HMC에 연결된 노드 정보를 보여준다.
- ▶ lsppeer : CSM DB에 정의된 HMC device 및 관련 정보를 보여준다.
- ▶ lsrhws : cluster hardware resource 및 resource class를 보여

준다.

- ▶ `mkrhws` : cluster hardware resource class instance를 정의한다.
- ▶ `netboot` : cluster VLAN을 통하여 AIX node를 설치한다.
- ▶ `rconsole` : 원격 콘솔을 연다.
- ▶ `rpower` : 노드의 파워정보를 검색하거나, 해당노드를 `reset`, `power on/off` 시킨다.
- ▶ `systemid` : 원격으로 하드웨어를 접속하기 위해 필요한 ID/password를 세팅한다.

### 1.3.1 Security

Hardware control resource class는 `/var/ct/cfg/ctrmc.acls` 파일을 통해 접근이 제어된다. 기본적으로 `root user`만 하드웨어 제어 명령을 실행할 수 있으며, 나머지 유저는 `read-only access`만 가능하다. 특정 `user`에 대해서 실행 권한을 줄 수도 있지만, 여기서는 생략하기로 한다.



## 2. CSM 설치 준비

### 2.1 Cluster Hardware 준비

CSM 클러스터 내에서는 512개까지의 System p 노드를 구성가능하며, 하나의 OS image가 하나의 노드로 간주된다. 현재 3호기에서 사용 중인 Parallel Systems Support Program (PSSP)은 클러스터 내에 설치되면 안되고, PSSP와 CSM은 연동될 수도 없다.

#### 2.1.1 CSM Management Server 준비

Management Server를 구성하기 위해 필요한 사항들을 살펴보면 아래와 같다:

- ▶ 반드시 CD-ROM, DVD-ROM, 또는 DVD-RAM 드라이브가 있어야 한다.
- ▶ AIX 5L V5.2 이상을 지원하는 모델이어야 한다.
- ▶ 외부망까지 사용하기 위해서는 적어도 3개 이상의 네트워크 어댑터가 있어야 한다.

#### 2.1.2 CSM Managed Node 준비

Managed Node는 크게 제약사항이 없으므로 생략하기로 한다.

## 2.2 Cluster Software 준비

CSM을 설치하기 위해서는 CSM 소프트웨어와 연관된 모든 소프트웨어를 반드시 설치해야한다.

설치에 필요한 소프트웨어는 AIX 5L, CSM, Linux product media에서 찾을 수 있으며, CSM 소프트웨어 설치를 마치면, /opt/csm/README/에서 README 파일을 찾을 수 있다.

설치에 관련된 기본 정보 및 소프트웨어를 살펴보면 아래와 같다:

- ▶ Management Server
  - AIX 5L 5.3 Technology Level 5 (5300-05), or later
  
- ▶ AIX 5L node
  - AIX 5L V5.3 with Recommended Maintenance Package 5300-03, or later
  
- ▶ CSM 1.6
  - AIX 5L 5.3 requires RSCT 2.4.6.0, or later
  
- ▶ CSM for AIX 5L software
  - csm.core
  - csm.dsh
  - csm.diagnostics

- csm.deploy
  - csm.server
  - csm.hc\_utils
  - csm.client
  - csm.msg
  - csm.hpsnm
  - csm.gui.dcem
  - csm.ll (optional)
  - csm.pe (optional)
  - csm.pessl (optional)
  - csm.essl (optional)
  - csm.gpfs (optional)
- ▶ AIX 5L open source software
- conserver-8.1
  - tcl-8.3.3-1
  - tk-8.3.3-1
  - expect-5.32-1
  - openCIMOM-0.8-1
  - openssl
- ▶ OpenSSH and prerequisite software
- openssh.base
  - openssh.license
  - openssh.man.en\_US
  - openssh.msg.en\_US
  - openssh.msg.EN\_US
  - NAS 1.4

- ▶ For the AIX 5L Kerberos server
  - krb5.server
  - krb5.msg.lang
  - krb5.client
  - krb5.doc.lang
  
- ▶ For the AIX 5L Kerberos client nodes
  - krb5.client
  - krb5.msg.lang

### 3. CSM 설치

#### 3.1 Node hostname 정의

CSM 클러스터에 포함될 각 node의 hostname을 /etc/hosts 파일에 정의한다.

```
150.183.143.41 csm1
10.10.10.101 csm
##### Cluster NetWork #####
10.10.10.1 clsn01
10.10.10.2 clsn02
10.10.10.3 clsn03
.
##### Public NetWork #####
150.183.143.31 node01
150.183.143.32 node02
150.183.143.33 node03
.
##### HMC Mgmt NetWork #####
11.11.11.101 csm_mgmt
11.11.11.2 HMCp595
#####SNI config#####
20.20.10.1 cl01s0
20.20.10.2 cl02s0
20.20.10.3 cl03s0
.
.
.
```

### 3.2 PATH 및 MANPATH 설정

CSM 관련 PATH 및 MANPATH를 설정한다.

```
export PATH=$PATH:/opt/csm/bin
export MANPATH=$MANPATH:/opt/csm/man
```

### 3.3 설치관련 파일시스템 생성

CSM 설치에 필요한 기본적인 정보가 저장되는 /csminstall 파일시스템과 NIM 관련 정보가 저장되는 /export/nim 파일시스템을 생성한다.

```
crfs -v jfs2 -g rootvg -m /csminstall -a size=4096M -a bf=true
crfs -v jfs2 -g rootvg -m /export/nim -a size=4096M -a bf=true
```

### 3.4 클러스터 소프트웨어 설치

“2.2 클러스터 소프트웨어 준비”에서 열거한 클러스터 소프트웨어를 설치한다. 해당 소프트웨어는 AIX 5L CD, Expansion Pack, CSM CD, Linux toolbox CD에서 찾을 수 있다.

설치가 완료되면, 해당 소프트웨어의 update (PTF)도 설치한다.

### 3.5 Remote Shell 속성 설정

CSM 클러스터 내에서 사용할 Remote Shell의 속성을 정의한다. CSM에서는 dsh을 remote shell로 사용하며, dsh을 이용하여 management server에서 각 node로 remote command를 수행할 수 있다.

Remote Shell은 rsh 및 ssh 중에서 설정가능하며, 4호기는 rsh로 정의하였다.

```
csmconfig RemoteShell=/usr/bin/rsh
```

일시적으로 Shell을 바꾸기 위해서는 아래와 같이 하면 된다.

```
export DSH_REMOTE_CMD=/usr/bin/ssh
```

### 3.6 License 인증

아래와 같이 CSM license를 등록할 수 있으며, 동시에 CSM 구성에 필요한 기본 파일들을 /csminstall 디렉토리로 복사할 수 있다.

```
csmconfig -c -L
```

### 3.7 Hardware Control Point UserID and Password

Hardware Control을 사용하기 위해서는 hardware control point user ID 및 password를 설정해야 한다. HMC를 통하여 remote hardware를 제어하기 위해서는 systemid 명령어를 통하여, user ID 및 password를 CSM에 등록해야한다. password는 HMC의 hscroot user의 password인 "abc123"을 사용한다.

```
systemid HMCp595 hscroot
```



### 3.8 CRHS Setup

HPS를 사용하기 위해 구성하는 CRHS는 installation server인 management server setup이후에 구성해도 상관없다. 즉, 기본적인 CSM구성과는 독립적이라 할 수 있다.

- ▶ IBM Hardware Management Consoles (HMCs)는 일반적으로 두개의 ethernet port를 가진다.  
CRHS service network (management network)은 반드시 eth0를 사용해야 한다.
- ▶ HMC 소프트웨어 upgrade 시, 반드시 CRHS를 재구성해야 한다.
- ▶ "hdwr\_svr" 는 system p 노드를 관리하기 위해서는 꼭 필요한 daemon이며, Bulk Power Assembly (BPA)와 Service Processor (SP)와 통신하기 위해, management server와 HMC에서 구동된다.
- ▶ "addpeer" 명령어는 cluster peer domain의 HMC상에서 구동되는 hardware server daemon을 cluster mode로 구동하게 한다. 기존에 HMC상에서 standalone mode로 구동되고 있었기 때문에 반드시 HMC를 reboot 해야 한다.
- ▶ CRHS에서 각 node의 hardware정보를 수집하기 위해서는 HMC의 eth0 SLP port (427)을 활성화 시켜야 한다. 이는 eth0 firewall window 상에서 변경가능하다.

### 3.8.1 Setup DHCP server on Management Server

DHCP를 구성하기 위해서는 반드시 `bos.net.tcp.server fileset`이 설치되어있어야 한다.

#### 1) bootp process 비활성화

- DHCP server와 NIM master 기능을 DHCP가 수행하게 만들기 위해 `bootp daemon`을 정지시킨다.

```
ps -ef | grep bootp  
kill the bootp process
```

#### 2) /etc/inetd.conf 파일 수정

- reboot 시, `bootp daemon`이 구동되는 것을 막기 위해, 아래와 같이 주석처리 한다.

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

#### 3) refresh inetd service

```
refresh -s inetd
```

#### 4) /etc/dhcpd.conf 파일 수정

```
leaseTimeDefault      0xffffffff
leaseExpireInterval   7 year
supportBOOTP          yes
supportUnlistedClients yes

network 11.11.11.0 255.255.255.0
{ #This network is used for SPs, BPAs and HMCs
    option 51      0xffffffff
    hostnamepolicy suggested
    subnet 11.11.11.0      11.11.11.2-11.11.11.254
}
```

### 5) DHCP server 구동 및 확인

- dhcpd daemon을 구동시키고 아래와 같이 확인한다.
- 처음 설치 시, HMC만 power가 켜져 있는 상태이므로 HMC에 첫 번째 IP가 할당된다.

```
#startsrc -s dhcpd
#lssrc -ls dhcpd
Log File:                /usr/tmp/dhcpd.log
Log Level:                0x806
Client Expire Interval:   3600
Reserve Expire Interval:  900
Bad Addr Reclaim Interval: 4294967295
Database Save Interval:   3600
IP Address      Status   Duration  Time Stamp  Client ID
-----
11.11.11.2     Leased  Infinite  Oct 16 04:39 1-00145ec75978
11.11.11.3     Free
11.11.11.4     Free
11.11.11.5     Free
.
.
```

## 6) 595 Frame power-on

- 순차적으로 IP를 할당시키기 위해서는 각 frame을 순차적으로 power-on 시킨다. 이때, 595 각 frame에 4개의 IP가 할당되는 것을 "lssrc -ls dhcpcsd" 명령어로 확인한다.
- 현재 구성된 정보를 보면 아래와 같다.

Log File:	/usr/tmp/dhcpcsd.log				
Log Level:	0x806				
Client Expire Interval:	3600				
Reserve Expire Interval:	900				
Bad Addr Reclaim Interval:	4294967295				
Database Save Interval:	3600				
IP Address	Status	Duration	Time Stamp	Client ID	
-----	-----	-----	-----		
11.11.11.2	Leased	Infinite	Oct 16 04:39	1-00145ec75978	
11.11.11.3	Leased	Infinite	Oct 10 18:22	1-00096bad76da	
11.11.11.4	Leased	Infinite	Oct 10 18:23	1-00096bad6dcb	
11.11.11.5	Leased	Infinite	Oct 11 22:43	1-00096bad2b54	
11.11.11.6	Leased	Infinite	Oct 11 22:43	1-00096bad2b52	
11.11.11.7	Leased	Infinite	Oct 10 18:28	1-00096badc4f6	
11.11.11.8	Leased	Infinite	Oct 10 18:28	1-00096badc467	
.					
.					

## 7) Hardware Element 탐색

- Hardware Server Discovery Agent를 사용하여 각각의 hardware element를 탐색한다. 탐색 정보는 share repository에 저장되며, 명령어 결과는 아래와 같다.
- IP가 595 각 프레임별 4개, 51A에 1개씩 할당되어있음을 확인할 수 있다.

```
# hwsda -a
```

device	type-model	serial number	IP addresses	
BPA	9458-100	91806B0	10.0.0.230	11.11.11.28
BPA	9458-100	918060H	10.0.0.236	11.11.11.24
BPA	9458-100	91806AM	10.0.0.237	11.11.11.40
BPA	9458-100	91806BD	10.0.0.242	11.11.11.32
BPA	9458-100	91806AC	10.0.0.243	11.11.11.20
BPA	9458-100	91805YV	10.0.0.244	11.11.11.12
BPA	9458-100	91806AZ	10.0.0.247	11.11.11.36
BPA	9458-100	91806AR	10.0.0.249	11.11.11.16
BPA	9458-100	91805Y1	10.0.0.250	11.11.11.8
BPA	9458-100	91805P6	10.0.0.254	11.11.11.4
HMC	7310C05	6533A2B	11.11.11.2	
BPA	9458-100	91805P6	11.11.11.3	10.0.0.253
SP	9119-595	832CD80	11.11.11.5	10.0.0.251
SP	9119-595	832CD80	11.11.11.6	10.0.0.252
BPA	9458-100	91805Y1	11.11.11.7	10.0.0.233
SP	9119-595	832CE90	11.11.11.9	10.0.0.216
SP	9119-595	832CE90	11.11.11.10	10.0.0.223
BPA	9458-100	91805YV	11.11.11.11	10.0.0.235
SP	9119-595	832CE60	11.11.11.13	10.0.0.248
SP	9119-595	832CE60	11.11.11.14	10.0.0.222
BPA	9458-100	91806AR	11.11.11.15	10.0.0.230
SP	9119-595	832CE30	11.11.11.17	10.0.0.225
SP	9119-595	832CE30	11.11.11.18	10.0.0.217
BPA	9458-100	91806AC	11.11.11.19	10.0.0.234
SP	9119-595	832CDA0	11.11.11.21	10.0.0.226
SP	9119-595	832CDA0	11.11.11.22	10.0.0.215
BPA	9458-100	918060H	11.11.11.23	10.0.0.239
SP	9119-595	832CE40	11.11.11.25	10.0.0.224
SP	9119-595	832CE40	11.11.11.26	10.0.0.227
BPA	9458-100	91806B0	11.11.11.27	10.0.0.229
SP	9119-595	832CE50	11.11.11.29	10.0.0.232
SP	9119-595	832CE50	11.11.11.30	10.0.0.218
BPA	9458-100	91806BD	11.11.11.31	10.0.0.221
SP	9119-595	832CEA0	11.11.11.33	10.0.0.231
SP	9119-595	832CEA0	11.11.11.34	10.0.0.240
BPA	9458-100	91806AZ	11.11.11.35	10.0.0.220
SP	9119-595	832CD90	11.11.11.37	10.0.0.246
SP	9119-595	832CD90	11.11.11.38	10.0.0.228
BPA	9458-100	91806AM	11.11.11.39	10.0.0.219
SP	9119-595	832CEB0	11.11.11.41	10.0.0.241
SP	9119-595	832CEB0	11.11.11.42	10.0.0.229
SP	9110-51A	062CCE0	11.11.11.43	192.168.3.147
SP	9110-51A	062CCC0	11.11.11.44	192.168.3.147
SP	9110-51A	062CCD0	11.11.11.45	192.168.3.147
SP	9110-51A	062CCF0	11.11.11.46	192.168.3.147

### 3.8.2 Setup HMC Cluster Peer Domain

Shared database를 사용하기 위해서는 HMC cluster peer domain을 구성해야 한다. "addpeer" 명령어로 특정 HMC 상에 shared database를 생성할 수 있으며, 이 정보를 바탕으로 HMC GUI 화면에서 각 Frame 및 Node 정보를 보는 것이 가능해진다. 즉, CRHS 구성과는 별도로 기존에 POWER5 시스템 노드 정보를 HMC에서 인식시키는 것과는 다른 방식으로, HMC 상에서 제어할 수 있는 하드웨어 정보가 HMC cluster peer domain을 구성해야지만 비로소 HMC 상에서 인식된다.

#### 1) "addpeer -v 11.11.11.2"

- "addpeer" 명령어는 아래와 같은 일을 수행하며, 반드시 제대로 수행되었는지 확인해야 한다.
- "addpeer" 명령어를 수행한 후에는 HMC를 반드시 reboot 해야 한다.
- ▶ CSM DB상에 non-node device로 정의한다.
- ▶ Management server에 의해 관리되도록 설정한다.
- ▶ HMC에서 peer domain을 구성한다.
- ▶ Peer domain의 HMC member를 생성한다.
- ▶ Peer domain을 활성화시킨다.

- ▶ Management server 상에서 "hdwr\_svr" daemon을 구동한다.
- ▶ HMC mode를 standalone에서 cluster로 변경한다.
- ▶ HMC에서 IBM.HWSVRRM daemon을 구동한다.
- ▶ HMC에서 rmc\_agent를 구동한다.
- ▶ Discovery agent를 polling mode로 변경한다.

## 2) hdwr\_svr daemon 확인

- "addpeer" 명령어가 제대로 수행되었다면, 아래와 같이 Management server에 "hdwr\_svr" daemon이 정상적으로 추가되었음을 확인할 수 있다. 만약, 문제가 있다면, /var/log/csm/addpeer.log를 참조한다.

```
#lssrc -s hdwr_svr
Subsystem      Group          PID           Status
hdwr_svr      hdwr_svr      372960       active
```

## 3) CSM DB 상에 정의된 device 확인

- HMC가 정상적으로 CSM DB상에 등록되었는지 아래와 같이 확인한다.

```
#lshwdev -a Name
11.11.11.2
#lshwdev -a Mode
11.11.11.2: Managed
```



#### 4) 각 Frame을 HMC에 할당

- 각각의 Frame 및 SP 정보를 확인한 후, 해당 HMC에 링크를 시킨다. 즉, Hardware Control Point를 각 Frame 및 node에 할당한다.

##### ▶ hardware element 확인

```
#lsrhws -e
-----
Element_Type      = "FSP"
Element_IP_A      = "11.11.11.6"
Element_IP_B      = "10.0.0.252"
Element_MTMS      = "9119-595*832CD80"
Element_CEC_Name  = " "
Element_FrameRH   = "0x2047 0xffff 0x57b3903b 0x75ab91e3 0x908c8776
0xc7c0c1c0"
Element_Slot      = "B"
Element_Frame_ID  = 1
Element_BPA_MTMS  = "9458-100*91805P6"
-----
Element_Type      = "BPA"
Element_IP_A      = "11.11.11.3"
Element_IP_B      = "10.0.0.253"
Element_MTMS      = "9458-100*91805P6"
Element_CEC_Name  = " "
Element_FrameRH   = "0x2047 0xffff 0x57b3903b 0x75ab91e3 0x908c8776
0xc7c0c1c0"
Element_Slot      = "A"
Element_Frame_ID  = 1
Element_BPA_MTMS  = "9458-100*91805P6"
.
.
```

##### ▶ hardware frame 확인

- Frame이 HMC에 연결되지 않았을 경우, Frame\_ManagerRH는 모두 '0'를 가진다.
- Link를 거쳐 정상적으로 Frame/Node가 할당되면, 각 Frame은 HMC의 Frame 값을 가진다.

```

#lsrhws -f
-----
Frame_ID          = 1
Frame_Type        = " "
Frame_ManagerRH_A = "0x0000 0x0000 0x00000000 0x00000000
0x00000000 0x00000000"
Frame_ManagerRH_B = "0x0000 0x0000 0x00000000 0x00000000
0x00000000 0x00000000"
Frame_BPA_MTMS    = "9458-100*91805P6"
-----
Frame_ID          = 2
Frame_Type        = " "
Frame_ManagerRH_A = "0x0000 0x0000 0x00000000 0x00000000
0x00000000 0x00000000"
Frame_ManagerRH_B = "0x0000 0x0000 0x00000000 0x00000000
0x00000000 0x00000000"
Frame_BPA_MTMS    = "9458-100*91805Y1"
.
.

```

▶ hardware manager 확인

```

#lsrhws -m
-----
Manager_Type      = "HMC"
Manager_IP_A      = "11.11.11.2"
Manager_IP_B      = ""
Manager_Name      = "HMCp595"
Manager_MTMS      = "7310C05*6533A2B"
Manager_Configured = 1
-----
Manager_Type      = "CSM_MS"
Manager_IP_A      = "11.11.11.101"
Manager_IP_B      = ""
Manager_Name      = "csm1"
Manager_MTMS      = ""
Manager_Configured = 0

```

- ▶ 각 Frame을 HMC에 할당
  - 위의 결과처럼 이상이 없을 경우, 아래와 같이 각 Frame을 HMC에 할당한다.

```
#chrhws -s 'Frame_BPA_MTMS=="9458-100*91805P6" -L
'Manager_MTMS=="7310C05*6533A2B"'
```

- ▶ hardware frame 확인
  - 제대로 Link가 되었다면, 아래와 같이 Frame\_ManagerRH가 바뀌었음을 확인할 수 있다.
  - 4호기의 경우, 동일 HMC에서 각 Frame/Node가 제어되기 때문에, 모든 Frame이 같은 값을 가짐을 확인할 수 있다.

```
#lsrhws -f
-----
Frame_ID          = 1
Frame_Type        = " "
Frame_ManagerRH_A = "0x2048 0xffff 0x57b3903b 0x75ab91e3
0x908c8776 0xcce8ee98"
Frame_ManagerRH_B = "0x0000 0x0000 0x00000000 0x00000000
0x00000000 0x00000000"
Frame_BPA_MTMS    = "9458-100*91805P6"
-----
Frame_ID          = 2
Frame_Type        = " "
Frame_ManagerRH_A = "0x2048 0xffff 0x57b3903b 0x75ab91e3
0x908c8776 0xcce8ee98"
Frame_ManagerRH_B = "0x0000 0x0000 0x00000000 0x00000000
0x00000000 0x00000000"
Frame_BPA_MTMS    = "9458-100*91805Y1"
.
.
```

- ▶ Setup Central Electronics Complex (CEC)/Bulk Power Assembly (BPA) password
  - 각 Frame/Node 접속 및 관리를 위해 HMC/general/admin user에 대한 password 정보를 CSM에 지정한다. Password는 각각 abc123/general/admin으로 지정한다.

```
#systemid -f -s IP_address HMC/general/admin
```

- ▶ Reboot HMC
  - 이 시점부터 HMC GUI 상에 각 Frame/CEC 정보가 나타난다.
- ▶ Frame number 변경
  - Frame number를 변경하기 위해서는 모든 Frame/CEC의 power를 off해야 한다.
  - "hwr\_svr" daemon이 활성화 되어있는지 확인한다.
  - 아래의 명령어 또는 HMC 상에서 변경가능하다.

```
#frame -m "9458-100*91805Y1" -i 2
```

- ▶ 각 Frame의 HMC 할당 정보 확인
  - 아래와 같이 각 Frame/Node가 HMC에 할당되었음을 확인할 수 있다.
  - 아래 정보는 1번 Frame에 대한 내용이며, Frame 정보, FSP와 BPA에 대한 Primary/Secondary Element 정보, Manager (HMC) 정보를 한꺼번에 보여준다.

```

#srhws -s 'Frame_BPA_MTMS=="9458-100*91805P6"' -L
-----
Frame_ID          = 1
Frame_Type        = " "
Frame_ManagerRH_A = "0x2048 0xffff 0x57b3903b 0x75ab91e3 0x908c8776 0xcce8ee98"
Frame_ManagerRH_B = "0x0000 0x0000 0x00000000 0x00000000 0x00000000 0x00000000"
Frame_BPA_MTMS    = "9458-100*91805P6"

Element_Type      = "FSP"
Element_IP_A      = "11.11.11.5"
Element_IP_B      = "10.0.0.251"
Element_MTMS      = "9119-595*832CD80"
Element_CEC_Name  = " "
Element_FrameRH   = "0x2047 0xffff 0x57b3903b 0x75ab91e3 0x908c8776 0xc7c0c1c0"
Element_Slot      = "A"
Element_Frame_ID  = 1
Element_BPA_MTMS  = "9458-100*91805P6"

Element_Type      = "FSP"
Element_IP_A      = "11.11.11.6"
Element_IP_B      = "10.0.0.252"
Element_MTMS      = "9119-595*832CD80"
Element_CEC_Name  = " "
Element_FrameRH   = "0x2047 0xffff 0x57b3903b 0x75ab91e3 0x908c8776 0xc7c0c1c0"
Element_Slot      = "B"
Element_Frame_ID  = 1
Element_BPA_MTMS  = "9458-100*91805P6"

Element_Type      = "BPA"
Element_IP_A      = "11.11.11.3"
Element_IP_B      = "10.0.0.253"
Element_MTMS      = "9458-100*91805P6"
Element_CEC_Name  = " "
Element_FrameRH   = "0x2047 0xffff 0x57b3903b 0x75ab91e3 0x908c8776 0xc7c0c1c0"
Element_Slot      = "A"
Element_Frame_ID  = 1
Element_BPA_MTMS  = "9458-100*91805P6"

Element_Type      = "BPA"
Element_IP_A      = "11.11.11.4"
Element_IP_B      = "10.0.0.254"
Element_MTMS      = "9458-100*91805P6"
Element_CEC_Name  = " "
Element_FrameRH   = "0x2047 0xffff 0x57b3903b 0x75ab91e3 0x908c8776 0xc7c0c1c0"
Element_Slot      = "B"
Element_Frame_ID  = 1
Element_BPA_MTMS  = "9458-100*91805P6"

Manager_Type      = "HMC"
Manager_IP_A      = "11.11.11.2"
Manager_IP_B      = ""
Manager_Name      = "HMCp595"
Manager_MTMS      = "7310C05*6533A2B"
Manager_Configured = 1

```

### 3.9 CSM 설치 확인

Management server가 제대로 설치되고 사용가능 여부를 확인하기 위해 `ibm.csm.ms probe`를 수행한다.

```
#probemgr -p ibm.csm.ms -l 0
Running probe /opt/diagnostics/probes/ibm.os.network.enabled.
Probe ibm.os.network.enabled was run successfully.
Running probe /opt/diagnostics/probes/ibm.os.network.hostname.
Probe ibm.os.network.hostname returned the following information.
ibm.os.network.hostname:trace:My hostname seems to be csm1.
Probe ibm.os.network.hostname was run successfully.
Running probe /opt/diagnostics/probes/ibm.os.network.ifaces.
Probe ibm.os.network.ifaces was run successfully.
Running probe /opt/diagnostics/probes/ibm.os.network.routes.
Probe ibm.os.network.routes was run successfully.
Running probe /opt/diagnostics/probes/ibm.os.network.ipforward.
Probe ibm.os.network.ipforward returned the following information.
ibm.os.network.ipforward:trace:IP forwarding seems to be enabled.
Probe ibm.os.network.ipforward was run successfully.
Running probe /opt/diagnostics/probes/ibm.os.network.ping.
Probe ibm.os.network.ping returned the following information.
ibm.os.network.ping:trace:ping -nqc10 -w10 10.10.10.101 .
ibm.os.network.ping:trace:ping -nqc10 -w10 11.11.11.101 .
ibm.os.network.ping:trace:ping -nqc10 -w10 150.183.143.41 .
ibm.os.network.ping:trace:ping -nqc10 -w10 127.0.0.1 .
Probe ibm.os.network.ping was run successfully.
Running probe /opt/diagnostics/probes/ibm.os.network.
Probe ibm.os.network returned the following information.
ibm.os.network:trace:The network probe has been run.
Probe ibm.os.network was run successfully.
Running probe /opt/diagnostics/probes/ibm.rmc.
Probe ibm.rmc returned the following information.
ibm.rmc:trace:Checking if the RMC daemons are running...
ibm.rmc:trace: Running cmd: LANG=C /usr/bin/lsrsrc -a
ibm.rmc:trace:Gathering information from the ctrmc daemon...
ibm.rmc:trace: Running cmd: LANG=C /usr/bin/lsrsrc -ls ctrmc
ibm.rmc:trace:Checking the RMC resource classes...
ibm.rmc:trace: Running cmd: /usr/bin/lsrsrc
ibm.rmc:trace:Checking the IBM.Host resource class...
ibm.rmc:trace: Running cmd: LANG=C /usr/bin/lsrsrc IBM.Host
Probe ibm.rmc was run successfully.
Running probe /opt/diagnostics/probes/ibm.csm.dmsrm.
Probe ibm.csm.dmsrm returned the following information.
ibm.csm.dmsrm:trace:Executing: /bin/lspp -l "csm.server"
ibm.csm.dmsrm:trace:Executing: lssrc -s IBM.DMSRM
ibm.csm.dmsrm:trace:DMSRM is active.
ibm.csm.dmsrm:trace:Executing: /opt/csm/bin/lsnode
Probe ibm.csm.dmsrm was run successfully.
Running probe /opt/diagnostics/probes/ibm.rmc.errm.
Probe ibm.rmc.errm returned the following information.
ibm.rmc.errm:trace:Executing: lssrc -s IBM.ERRM
ibm.rmc.errm:trace:ERRM is active.
ibm.rmc.errm:trace:Executing: /usr/sbin/rsct/bin/lscondition
Probe ibm.rmc.errm was run successfully.
Running probe /opt/diagnostics/probes/ibm.csm.predefined-conditions.
Probe ibm.csm.predefined-conditions returned the following information.
```

```

ibm.csm.predefined-conditions:trace:Checking that condition "AIXNodeCoreDump" is defined.
ibm.csm.predefined-conditions:trace:Checking that condition "AllServiceableHardwareEvents" is
defined.
ibm.csm.predefined-conditions:trace:Checking that condition "AllServiceableSwitchEvents" is
defined.
ibm.csm.predefined-conditions:trace:Checking that condition "AnyNodeAnyLoggedError" is
defined.
ibm.csm.predefined-conditions:trace:Checking that condition "AnyNodeFileSystemInodesUsed" is
defined.
.
.
.
Running probe /opt/diagnostics/probes/ibm.csm.predefined-nodegroups.
Probe ibm.csm.predefined-nodegroups was run successfully.
Running probe /opt/diagnostics/probes/ibm.csm.ms.
Probe ibm.csm.ms returned the following information.
ibm.csm.ms:trace:Checking directory /opt/csm ...
ibm.csm.ms:trace:/opt/csm exists.
ibm.csm.ms:trace:Checking directory /opt/csm/bin ...
ibm.csm.ms:trace:/opt/csm/bin exists.
ibm.csm.ms:trace:Checking directory /opt/csm/csmbin ...
ibm.csm.ms:trace:/opt/csm/csmbin exists.
.
.
.
ibm.csm.ms:trace:Checking for packages : csm.client*, csm.core*, csm.deploy*,
csm.diagnostics*, csm.dsh*, csm.hc_utils, csm.server*.
ibm.csm.ms:trace:Checking for packages : conserver-8.1.7, expect*, openCIMOM-0.8-1, tcl*,
tk*.
ibm.csm.ms:trace:Check if the CFM cronjob is enabled.
Probe ibm.csm.ms was run successfully.

```

### 3.10 Kerberos setup

PSSP에서 사용하는 Kerberos V4와는 다르게 csmconfig 명령어로 CSM상에서 Kerberos 세팅이 가능하다. 아래에 열거한 일들을 따로 혹은 동시에 처리할 수 있으며, 각각의 일들이 의미하는 숫자의 조합으로 가능하게 한다.

- ▶ Management상에 Kerberos server를 생성한다.
- ▶ Kerberos database에 CSM principal을 생성한다.
- ▶ Kerberos database에 CSM 클러스터내에 정의된 각 노드의 host principal을 생성한다.



- ▶ Managed node의 root 디렉토리에 .k5login 파일을 생성한다.

```
#csmconfig SetupKRB5=n (숫자 조합)
  0 No CSM setup
  1 configure server on MS
  2 configure client on managed node
  4 create CSM principal
  8 create host principal
```

예를 들어, SetupKRB5=15라고 지정하면 모든 Kerberos 구성요소들이 CSM 클러스터내에서 세팅된다.

### 3.11 Installation server setup

지금까지의 스텝들이 CSM 클러스터를 생성하기 위해 CSM Management server를 구성하는 단계였다면, 이제부터는 Managed Node를 구성하기 위하여 각 노드, device 정보를 정의하고 그 정보를 바탕으로 각 노드를 설치하는 과정이다.

KISTI 4호기 1차 시스템은 그 구성이 복잡하지 않아, 아래 몇 가지의 항목만 세팅하면 된다.

Installation server 구성을 위해 먼저 생각해야 되는 항목을 살펴보자:

- ▶ Hostname : clsn01 ~ clsn10

- ▶ ManagementServer : csm

==> 여기서부터는 주로 Cluster VLAN을 사용하므로 위의 호스트명은 전부 Cluster VLAN에 해당된다.

### 3.11.1 노드 정의

클러스터 노드를 정의하는 단계이며, CSM database상에 아래의 항목들을 정의한다.

하드웨어 종류에 따라 필요한 항목이 약간씩 바뀌며, 여기서는 HMC-attached system p 노드에 대해서만 작성하였다. 아래의 항목들은 NIM installation 전까지만 정의되면 되며, 언제든지 여러 가지 명령어로 정의 가능하다. 특히, Network 항목들은 "getadapters" 명령어로 나중에 입력하게 된다.

- ▶ Hardware control attributes  
(HMC-attached system p nodes)

PowerMethod : hmc HWControlPoint : hostname or IP of HMC HWControlNodeId : partition name of LPAR
---

- ▶ Installation attributes

InstallAdapterNetmask : on the same subnet, left blank InstallAdapterDuplex : auto/half/full InstallAdapterMacaddr <== getadapters command InstallAdapterSpeed : auto/10/100/1000 InstallAdapterType : ent InstallAdapterName : etho...
--

▶ Node installation attributes

```
InstallOSName : AIX
```

노드를 정의하는 방식은 직접 수동으로 입력하는 방식과 management server에서 remote power control point (HMC)를 통하여 수집된 정보를 이용하여 입력하는 방식이 있다. 여기서는 HMC를 통하여 탐색된 노드의 정보를 이용하는 방법만 다루기로 한다.

아래와 같이 "lshwinfo" 명령어를 이용하여 노드 정보를 수집할 수 있으며, 수집된 정보를 파일로 저장하여 "no\_hostname" 부분을 해당 hostname으로 바꿔준다.

```
#lshwinfo -p hmc -c HMCp595
#Hostname::PowerMethod::HWControlPoint::HWControlNodeID::LParID::HWType::
HWModel::HWSerialNum::DeviceType::UUID
no_hostname::hmc::HMCp595::83-2CD80::1::9119::595::832CD80:::
no_hostname::hmc::HMCp595::83-2CE90::1::9119::595::832CE90:::
no_hostname::hmc::HMCp595::83-2CE60::1::9119::595::832CE60:::
no_hostname::hmc::HMCp595::83-2CE30::1::9119::595::832CE30:::
no_hostname::hmc::HMCp595::83-2CDA0::1::9119::595::832CDA0:::
no_hostname::hmc::HMCp595::83-2CE40::1::9119::595::832CE40:::
no_hostname::hmc::HMCp595::83-2CE50::1::9119::595::832CE50:::
no_hostname::hmc::HMCp595::83-2CEA0::1::9119::595::832CEA0:::
no_hostname::hmc::HMCp595::83-2CD90::1::9119::595::832CD90:::
no_hostname::hmc::HMCp595::83-2CEB0::1::9119::595::832CEB0:::
#lshwinfo -p hmc -c HMCp595 -o /tmp/ibm/nodemap
```

위의 정보 수정

```
=====
clsn01::hmc::HMCp595::83-2CD80::1::9119::595::832CD80:::
clsn02::hmc::HMCp595::83-2CE90::1::9119::595::832CE90:::
clsn03::hmc::HMCp595::83-2CE60::1::9119::595::832CE60:::
clsn04::hmc::HMCp595::83-2CE30::1::9119::595::832CE30:::
clsn05::hmc::HMCp595::83-2CDA0::1::9119::595::832CDA0:::
clsn06::hmc::HMCp595::83-2CE40::1::9119::595::832CE40:::
clsn07::hmc::HMCp595::83-2CE50::1::9119::595::832CE50:::
clsn08::hmc::HMCp595::83-2CEA0::1::9119::595::832CEA0:::
clsn09::hmc::HMCp595::83-2CD90::1::9119::595::832CD90:::
clsn10::hmc::HMCp595::83-2CEB0::1::9119::595::832CEB0:::
```

마지막으로 HMC를 통해 탐색된 노드 정보를 이용하여 CSM DB에 각 항목을 정의한다.

```
#definenode      -M      /tmp/ibm/nodemap      InstallCSMVersion=1.6.0.10
InstallOSName=AIX
```

### 3.11.2 Device 정의

이미 "addpeer" 명령어를 통하여 HMC는 device로 CSM DB 상에 정의되어 있다. 이외에도 network switch, I/O unit등을 CSM DB상에 정의할 수 있다.

### 3.12 NIM Installation

전 단계에서 정의된 노드 정보를 바탕으로, CSM Management server에 NIM master를 구성하고 각 노드를 설치한다. 설치가 끝나면 설치된 노드는 post-script를 통해 자동으로 CSM 클러스터에 포함된다. 여기서는 설치의 편의를 위해 mksysb를 이용한 NIM 설치를 하였다.

### 3.12.1 각 노드 정의 확인

CSM DB상에 구성된 노드 정보를 확인한다.

```
#lsnode
clsn01
clsn02
clsn03
clsn04
clsn05
clsn06
clsn07
clsn08
clsn09
clsn10
#lsnode -l clsn01
  Hostname = clsn01
  AdapterStanzaFile =
  AllowManageRequest = 0 (no)
  CSMVersion = 1.6.0.10
  ChangedAttributes = {LastCFMUpdateTime}
  ConfigChanged = 0 (no)
  ConsoleMethod = hmc
  ConsolePortNum = 83-2CD80
  ConsoleRedirectionAfterPOST = 0
  ConsoleSerialDevice = ttyS0
  ConsoleSerialSpeed = 9600
  ConsoleServerName = HMCp595
  ConsoleServerNumber =
  FWSvcProc =
  FWSysBIOS =
  HWControlNodeId = 83-2CD80
  HWControlPoint = HMCp595
  HWModel = 595
  HWSerialNum = 832CD80
  HWType = 9119
  .
  .
```

### 3.12.2 Remote Power Control Check

CSM Management server에서 아래의 명령어를 사용하여 remote power control이 가능한지 확인한다.

```
#rpower -a query
clsn09 off
clsn03 off
clsn04 off
clsn01 off
clsn06 off
clsn05 off
clsn10 off
clsn08 off
clsn02 off
clsn07 off
```

### 3.12.3 Network Adapter 정보 수집

CSM Management server에서 "getadapters" 명령어로 NIM 설치에 필요한 network adapter의 정보를 수집한다. 반드시 시스템은 partition standby 상태여야하며, 처음 노드를 설치할 경우 시스템의 'openfirmware'와 연동하여 mac\_address 및 기본 정보들을 가져오게 된다. 정보를 파일에 저장한 후, 다시 "getadapters" 명령을 사용하여 CSM DB에 network 정보를 정의한다.(-w 옵션)

아래 내용은 순차적으로 CSM DB상에 정의하는 과정을 보여준다.

- ▶ "getadapters" 명령을 사용하여 파일에 저장 및 내용을 보는 예이다. 여기서 Duplex 및 Speed는 auto로 지정하였다.

```
#getadapters -t ent -D -s auto -d auto -n clsn01,clsn02,.. -z
/tmp/ibm/mystanza
#cat /tmp/ibm/mystanza
###CSM_ADAPTERS_STANZA_FILE###--do not remove this line
#---Stanza Summary-----
#   Date: Sun Oct 14 22:05:07 KORST 2007
#   Stanzas Added: 0
#---End Of Summary-----
clsn01::ent::00145ec7d502:::auto::auto
clsn02::ent::00145ec78fce:::auto::auto
clsn03::ent::00145ec745d8:::auto::auto
clsn04::ent::00145ec788c2:::auto::auto
clsn05::ent::00145ec78a0e:::auto::auto
clsn06::ent::00145ec7564c:::auto::auto
clsn07::ent::00145ec74124:::auto::auto
clsn08::ent::00145ec78fb8:::auto::auto
clsn09::ent::00145ec7a3e2:::auto::auto
clsn10::ent::00145ec79ff0:::auto::auto
```

- ▶ CSM DB에 가져온 정보를 정의한다.

```
#getadapters -w -f /tmp/ibm/mystanza
```

### 3.12.4 기본 NIM setup

여기서는 "nim\_master\_setup" 명령어만 알아보고 NIM에 관련된 자세한 설명은 생략한다.

CSM management server 상에 network interface 가 다수 존재하고 이미 구성을 마친 경우에는 "nim\_master\_setup -B" 명령어가 Cluster VLAN을 사용하지 않고 NIM 구성을 할 수 있기 때문에 먼저 NIM master\_net을 정의하는게 좋다. "nim\_master\_setup -B" 명령어를 사용하여 management server상에서 NIM을 구성하며, 이 명령어는 설치에 필요한 기본 NIM resource (예:spot, lpp\_source)를 자동으로 생성한다. "-B" 옵션은 CSM management server의 mksysb 이미지를 생성하지 못하게 한다. 반드시 AIX 5.3 설치 CD를 명령어 수행전에 CDRROM에 넣어주어야 한다.

### 3.12.5 NIM mksysb resource 생성

이미 하나의 노드를 설치한 후 만든 mksysb 이미지로 각 노드를 설치할 것이므로, "nim\_master\_setup -B" 명령어로 생성한 NIM resource외에도 mksysb resource가 필요하다. 아래 예는 /image/clsmksysb 이미지를 "cls\_mksysb"라는 이름의 NIM mksysb resource로 등록하는 것을 보여준다.

```
#nim -o define -t mksysb -a server=master -a location=/image/clsmksysb  
cls_mksysb
```



### 3.12.6 NIM machine 정의

NIM 설치 시 사용할 노드명을 CSM에 등록된 노드명과 일치시키기 위해서 "csm2nimnodes" 명령어를 이용한다. 이 명령어를 사용하면 CSM DB에 저장된 정보를 읽어 와서 NIM client를 정의하게 된다. 편의성을 위해 노드 그룹을 지정하였으며, "-N" 옵션으로 사용가능하다.

```
#nodegrp -a clsn01,clsn02,clsn03,.....,clsn08,clsn09,clsn10 computing
#csm2nimnodes -N computing
#lsnim |grep machines
master          machines      master
clsn08          machines     standalone
clsn02          machines     standalone
clsn01          machines     standalone
clsn04          machines     standalone
clsn09          machines     standalone
clsn07          machines     standalone
clsn06          machines     standalone
clsn05          machines     standalone
clsn03          machines     standalone
clsn10          machines     standalone
```

노드뿐만 아니라 CSM 그룹도 변환이 가능하며 아래와 같이 설정할 수 있다.

```
#csm2nimgrps -N computing
```

### 3.12.7 AIX 5L 인증방식 선택

NIM 설치 시 **standard AIX 5L** 인증방식을 사용하기 때문에, 인증 방식에 반드시 포함되어야 한다.

만약, 포함되어 있지 않으면, 아래와 같이 수정한다.

```
#lsauthent  
Kerberos 5  
#chauthent -k5 -std  
#lsauthent  
Kerberos 5  
Standard Aix
```

### 3.12.8 클러스터내에 노드 추가 준비

NIM 설치 후 자동으로 노드를 CSM 클러스터에 추가시키기 위하여 **customization**을 위한 **post script**를 NIM script resource로 등록하는 단계이다. "**csmsetupnim**" 명령어를 사용하며, 이 명령어는 노드 구성정보 파일을 생성하고 **script**를 NIM resource로 등록한다. 노드를 새로 설치하거나 재설치하는 경우 모두 이 명령어를 수행해야 한다.

```
#csmsetupnim -N computing
```

### 3.12.9 노드 bos\_inst operation

최종적으로 NIM 설치를 위하여 설치될 전체 노드를 설치를 위한 상태로 전환한다.

```
#nim -o bos_inst -a source=mksysb -a mksysb=cls_mksysb -a  
spot=530spot_res -a boot_client=no computing
```

### 3.12.10 노드 설치

노드 설치를 위한 모든 준비가 끝났으며, "netboot" 명령어로 network installation을 시작한다.

```
#netboot -N computing
```

설치 중에 아래와 같이 remote console을 열어서 설치를 모니터링 가능하다.

```
#rconsole -r -t -n clsn01
```

### 3.12.11 설치 후 점검

노드 설치가 끝나면 저절로 reboot이 되며, reboot후에 post script가 수행된다. 제대로 script가 수행되면 아래와 같은 결과들을 얻을 수 있다.

```
#lsnode -a Mode
clsn01: Managed
clsn02: Managed
clsn03: Managed
clsn04: Managed
clsn05: Managed
clsn06: Managed
clsn07: Managed
clsn08: Managed
clsn09: Managed
clsn10: Managed
#dsh -a date (Keberos 동작 확인)
clsn02: Tue Nov 27 02:26:58 KORST 2007
clsn04: Tue Nov 27 02:26:42 KORST 2007
clsn09: Tue Nov 27 02:26:45 KORST 2007
clsn07: Tue Nov 27 02:26:59 KORST 2007
clsn10: Tue Nov 27 02:26:38 KORST 2007
clsn03: Tue Nov 27 02:26:39 KORST 2007
clsn05: Tue Nov 27 02:26:55 KORST 2007
clsn06: Tue Nov 27 02:26:57 KORST 2007
clsn08: Tue Nov 27 02:26:48 KORST 2007
clsn01: Tue Nov 27 02:26:37 KORST 2007
```

만약, 'Managed'가 아닌 'Pre-managed'라는 결과가 나오는 노드가 있거나 "dsh" 명령이 제대로 수행되지 않는 노드가 있으면 post script를 따로 돌릴 수도 있다.

```
#updatenode -k -n clsn01
```

-k : remote command를 위해 Kerberos 정보를 갱신함

I,F,S 옵션은 default로 수행됨

=====

-I : customization script를 수행

-F : /cfmroot 내의 파일들을 노드로 copy

-S : /csminstall 내의 Linux RPM 패키지를 설치 또는 갱신함

\*\* 각각의 옵션이 명령어와 함께 명시되면 해당 옵션만 수행됨

## 4. HPS setup

High Performance Switch (HPS)는 기존의 3호기에서 사용하는 SP Switch2 topology를 그대로 사용한다. HPS를 사용하는 가장 큰 목적은 RAS 외에도 bandwidth 향상과 latency를 줄이는데 있다.

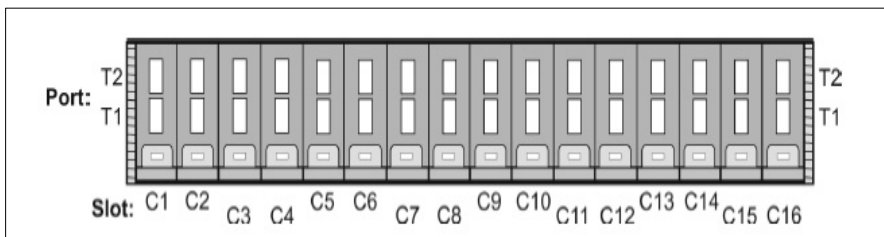
### 4.1 고려 사항

아래의 사항들이 반드시 만족되어야 한다.

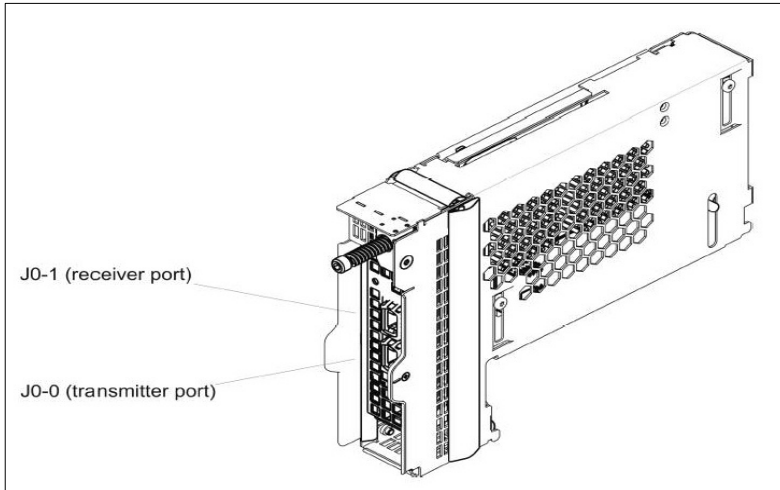
- ▶ csm.gui.websm, csm.hpsnm fileset 설치
- ▶ /var 파일 시스템 사이즈 : 1GB 이상
- ▶ CRHS setup
- ▶ AIX 5.2 or later
- ▶ AIX 5.3 requires RSCT Version 2.4.2.0 or later
- ▶ CSM 1.4.1 or later

### 4.2 용어 설명

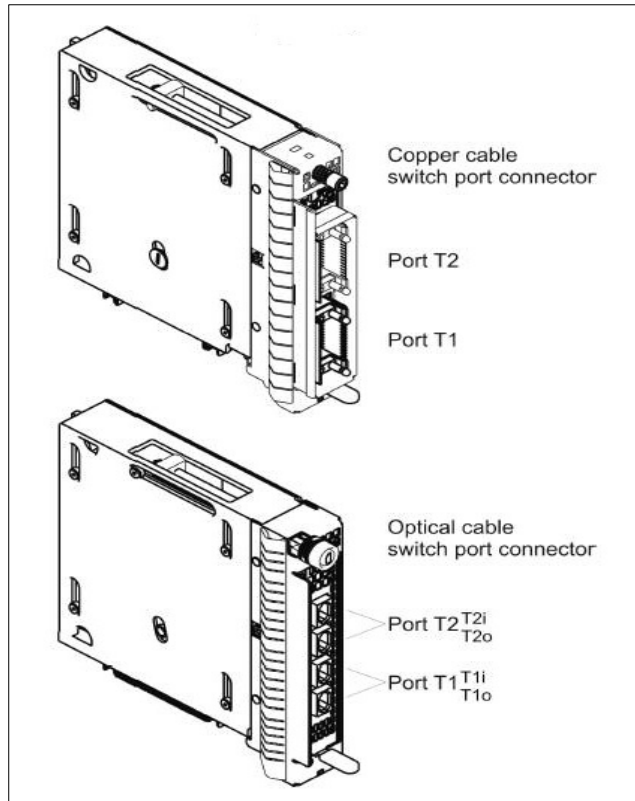
- ▶ Node Switch Board (NSB) : Switch drawer를 의미하며, 3개의 NSB가 설치되었다.



- ▶ Switch Network Interface (SNI) : 각 노드에 장착되는 Switch adapter



- ▶ Switch Port Connection Card (SPC) : NSB상에 꽂히는 Switch card T1, T2 port를 가지며 카드 당 2-link 가 형성 된다.



- ▶ Switch Port Slot : NSB 상에 Switch card가 꽂히는 공간 (C1~C16)
- ▶ Switch-to-Switch communication : NSB간의 통신을 의미하며 Switch Port Slot (1,2),(5,6),(9,10),(13,14) slot의 SPC 카드가 해당된다.
- ▶ Server-to-Switch communication : NSB와 노드간의 통신을 의미하며 SNI와 SPC를 연결한다. (3,4),(7,8),(11,12),(15,16) slot의 SPC 카드가 해당된다.

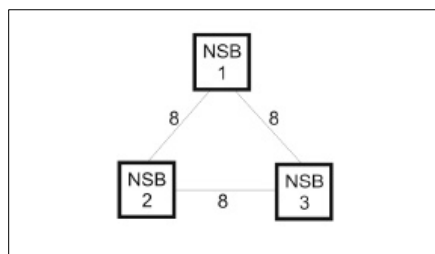


- ▶ Single network system : 모든 link가 통신이 가능한 network system으로 p595 HPS 구성이 여기에 해당된다.
- ▶ Two(dual) network system : 두개의 link pair가 서로 통신이 가능하지 않은 NSB로 연결되는 방식으로 각각의 network은 서로 Switch-to-Switch communication을 위한 연결을 하지 않으며, p575 HPS 구성이 여기에 해당된다.

### 4.3 KISTI HPS 구성

KISTI 4호기 시스템은 노드 당 4개 Link가 필요하기 때문에 전체 Server-to-Switch communication을 위해서 20개의 SPC 카드가 필요하다. 따라서 3개의 NSB가 single network으로 구성된다.

- ▶ NSB 간 연결  
 ==> 아래 그림과 같이 인접한 NSB로 Switch-to-Switch communication을 위한 8개의 port를 사용하여 연결된다.



## Switch-to-Switch Cabling Table

NSB 1	NSB 2	NSB 3	Cable serial number
Slot C1 Port T1		Slot C2 Port T1	
Slot C1 Port T2		Slot C2 Port T2	
Slot C2 Port T1	Slot C1 Port T1		
Slot C2 Port T2	Slot C1 Port T2		
Slot C5 Port T1		Slot C6 Port T1	
Slot C5 Port T2		Slot C6 Port T2	
Slot C6 Port T1	Slot C5 Port T1		
Slot C6 Port T2	Slot C5 Port T2		
Slot C9 Port T1		Slot C10 Port T1	
Slot C9 Port T2		Slot C10 Port T2	
Slot C10 Port T1	Slot C9 Port T1		
Slot C10 Port T2	Slot C9 Port T2		
Slot C13 Port T1		Slot C14 Port T1	
Slot C13 Port T2		Slot C14 Port T2	
Slot C14 Port T1	Slot C13 Port T1		
Slot C14 Port T2	Slot C13 Port T2		
	Slot C2 Port T1	Slot C1 Port T1	
	Slot C2 Port T2	Slot C1 Port T2	
	Slot C6 Port T1	Slot C5 Port T1	
	Slot C6 Port T2	Slot C5 Port T2	
	Slot C10 Port T1	Slot C9 Port T1	
	Slot C10 Port T2	Slot C9 Port T2	
	Slot C14 Port T1	Slot C13 Port T1	
	Slot C14 Port T2	Slot C13 Port T2	

► Server-to-Switch Communication

Server-to-Switch Connection Table

server number (N)	SNI location (S1 ~ S2)	SNI port number	SNI link address	Connection order	Corresponding switch port location	Switch cable serial number
N1	S1	Q1	N1S1Q1	1	S1C3T1	
N1	S1	Q2	N1S1Q2	1	S2C3T1	
N1	S2	Q1	N1S2Q1	11	S1C12T1	
N1	S2	Q2	N1S2Q2	11	S2C12T1	
N2	S1	Q1	N2S1Q1	2	S1C3T2	
N2	S1	Q2	N2S1Q2	2	S2C3T2	
N2	S2	Q1	N2S2Q1	12	S1C12T2	
N2	S2	Q2	N2S2Q2	12	S2C12T2	
N3	S1	Q1	N3S1Q1	3	S1C4T1	
N3	S1	Q2	N3S1Q2	3	S2C4T1	
N3	S2	Q1	N3S2Q1	13	S1C15T1	
N3	S2	Q2	N3S2Q2	13	S2C15T1	
N4	S1	Q1	N4S1Q1	4	S1C4T2	
N4	S1	Q2	N4S1Q2	4	S2C4T2	
N4	S2	Q1	N4S2Q1	14	S1C15T2	
N4	S2	Q2	N4S2Q2	14	S2C15T2	
N5	S1	Q1	N5S1Q1	5	S1C7T1	
N5	S1	Q2	N5S1Q2	5	S2C7T1	
N5	S2	Q1	N5S2Q1	15	S1C16T1	
N5	S2	Q2	N5S2Q2	15	S2C16T1	
N6	S1	Q1	N6S1Q1	6	S1C7T2	
N6	S1	Q2	N6S1Q2	6	S2C7T2	
N6	S2	Q1	N6S2Q1	16	S1C16T2	
N6	S2	Q2	N6S2Q2	16	S2C16T2	
N7	S1	Q1	N7S1Q1	7	S1C8T1	
N7	S1	Q2	N7S1Q2	7	S2C8T1	
N7	S2	Q1	N7S2Q1	17	S3C3T1	
N7	S2	Q2	N7S2Q2	21	S3C11T1	
N8	S1	Q1	N8S1Q1	8	S1C8T2	
N8	S1	Q2	N8S1Q2	8	S2C8T2	
N8	S2	Q1	N8S2Q1	18	S3C3T2	
N8	S2	Q2	N8S2Q2	22	S3C11T2	
N9	S1	Q1	N9S1Q1	9	S1C11T1	
N9	S1	Q2	N9S1Q2	9	S2C11T1	
N9	S2	Q1	N9S2Q1	19	S3C7T1	
N9	S2	Q2	N9S2Q2	23	S3C15T1	
N10	S1	Q1	N10S1Q1	10	S1C11T2	
N10	S1	Q2	N10S1Q2	10	S2C11T2	
N10	S2	Q1	N10S2Q1	20	S3C7T2	
N10	S2	Q2	N10S2Q2	24	S3C15T2	

## 4.4 HPS Network setup

물리적인 구성을 끝내고 나면 응용프로그램에서 HPS를 사용할 수 있도록 HPS network을 구성해야 한다.

### 4.4.1 HPS service login ID 생성

HPS network을 생성하고 관리하기 위해 CSM에서 HPS service login ID를 생성해야 한다.

"/opt/csm/hpsnm/tools/bin/setup\_service" script를 수행하면 "hpssvc" 라는 이름의 ID가 생성된다. "setup\_service" script 가 수행하는 일들을 살펴보면 아래와 같다.

- ▶ hpssvc group 생성
  - ▶ hpssvc user 생성
  - ▶ HPS Network Manager (HPSNM)를 수행할 수 있는 권한 부여
- ==> 한 가지 주의할 점은 생성 후 `account_locked=false`, `logintimes='blank'` 로 user profile을 바꿔주어야 아래 단계에서 CSM management server가 제대로 HMC에 추가된다.

#### 4.4.2 HMC에 CMS management server 추가

HMC에서 HPSNM을 관리하기 위해 HMC GUI상에서 CSM management server를 추가한다. 이때, user ID는 hpssvc를 사용한다. 정상적으로 HOST 추가작업이 끝나게 되면 기존에 HMC가 가지고 있던 노드 관리정보 외에도 HPSNM 관리 메뉴를 확인할 수 있다.

#### 4.4.3 HPSNM 시범 구동

정상적으로 HPS가 설치되었는지 확인하기 위해 테스트 과정을 거치는 단계이다. 임시적으로 HPS를 구동시켜서 topology 및 하드웨어의 이상 유무를 확인하는 단계라고 보면 된다.

- ▶ 모든 CEC을 power off 한다. 모든 Frame의 UEPO power 만 올려둔다.
- ▶ HPSNM GUI 화면에서 HPS Network Manager 선택
  - => Display Cluster Componets 선택
    - : 모든 NSB 와 Frame이 제대로 보이는지 확인
    - : 각 Frame number가 CRHS 구성 시 입력한 Frame number와 일치하는지 확인
  - => Select Logical Topology
    - : 3NSB\_0ISB\_48EP 선택
  - => Select HPSNM ELA Master
    - : 해당 HMC 선택
  - => Enable HPSNM Software for Switch Network Verification 선택

- : Enable HPSNM Software for Normal Operation을 선택하지 않도록 주의한다.
- : HPSNM을 구동시키거나 정지 후에는 반드시 HPSNM GUI 화면을 '새로 고침'해야 한다.

=> Switch Topology View 선택

- : "Select Logical Topology" 단계에서 설정한 값이 제대로 보이는지 확인
- : 각 SPC 카드 및 SNI 카드의 link 상태가 정상인지 점검

- ▶ CEC 구동 후, SNI 카드 정보를 CSM DB에 갱신하고, 각 노드를 customization 한다.

```
#getadapters -m dsh -t sni -z /tmp/ibm/snistanza -n clsn01
```

==> /tmp/ibm/snistanza 파일을 아래와 같이 수정한다.

```
clsn05:  
  adapter_type=sni  
  interface_type=sn  
  interface_name=sn0  
  location=U787C.001.9180592-P2-C8  
  machine_type=secondary  
  netaddr=20.20.10.5  
  subnet_mask=255.255.255.0
```

```
clsn05:  
  adapter_type=sni  
  interface_type=sn  
  interface_name=sn1  
  location=U787C.001.9180592-P2-C9  
  machine_type=secondary  
  netaddr=20.20.20.5  
  subnet_mask=255.255.255.0
```

```
clsn05:  
  adapter_type=sni  
  interface_type=sn  
  interface_name=sn2  
  location=U787C.001.9180592-P3-C8  
  machine_type=secondary  
  netaddr=20.20.10.15  
  subnet_mask=255.255.255.0
```

```
clsn05:  
  adapter_type=sni  
  interface_type=sn  
  interface_name=sn3  
  location=U787C.001.9180592-P3-C9  
  machine_type=secondary  
  netaddr=20.20.20.15  
  subnet_mask=255.255.255.0
```

```
clsn05:  
  machine_type=secondary  
  adapter_type=mlt  
  interface_type=ml  
  interface_name=ml0  
  netaddr=20.20.50.5  
  subnet_mask=255.255.255.0
```

==> 수정된 /tmp/ibm/snistanza 파일을 이용하여 CSM DB를 갱신하고, 각 노드를 customization 한다.

```
#getadapters -W -f snistanza05  
#updatenode -c -n clsn01
```

-c 옵션 : 각 노드의 secondary adapter를 구성한다.

- ▶ 각 노드를 reboot하면 SNI interface 정보를 제대로 가져오게 되는데, reboot 전에 Large Page 필요 수량을 확인하여 반영한다. Window 개수가 16당 109개 정도가 필요하므로 현재 window 개수가 64개로 설정되어있어 128\*4 개의 Large Page를 할당하였다.

```
node01:/usr/sni/aix53/debugtools> ./sni_calc_lpgg_cnt -a 4  
number_of_sni = 4  
num_windows = 16  
total_num_windows = 23  
spool_size = 0x2000000  
rpool_size = 0x2000000  
Total required LPs == 0x6d [109]  
  
node01:> dsh -n clsn01,clsn02 "echo y|vmo -r -o v_pinshm=1 -o  
lpgg_size=16777216  
-o lpgg_regions=128"  
  
csm1:> rpower -N computing reboot
```



#### 4.4.4 HPSNM 정상 구동

전체 노드의 HPS interface가 정상인지 확인 후, HPSNM을 구동 정지 시키고 이번에는 테스트가 아닌 "enable HPSNM for normal operation"으로 정상적으로 HPS network을 구동시킨다.

#### 4.4.5 HPSNM 명령어

- ▶ ntblstatus : Switch Window 할당 상태를 보여준다.
- ▶ chswnm : HPSNM을 구동, 정지 혹은 verification 모드로 구동시킨다.
  - => -a 옵션 : 구동
  - => -d 옵션 : 정지
  - => -q 옵션 : 현재 상태 검색
  - => -v 옵션 : verification mode
- ▶ lsswendpt : 각 노드의 SNI 카드 상태 정보를 확인할 수 있다.
- ▶ lsswenvir : NSB의 상태 정보를 확인할 수 있다.
- ▶ lsswtopol : SNI와 SPC 카드 간의 link 상태를 보여준다.
- ▶ lsswcomp : HPSNM이 정지되어있어야 확인 가능하며, 클러스터내 HPS Frame 정보를 보여준다.