PNNL-13266

Installation and Configuration Guide for the IBM SP Molecular Science Computing Facility William R. Wiley Environmental Molecular Sciences Laboratory Pacific Northwest National Laboratory

Damian Hazen***, Scott Jackson, Larry Orcutt*, Bill Pitre, Michael Valley**, Ralph Wescott, Tim Witteveen

Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352

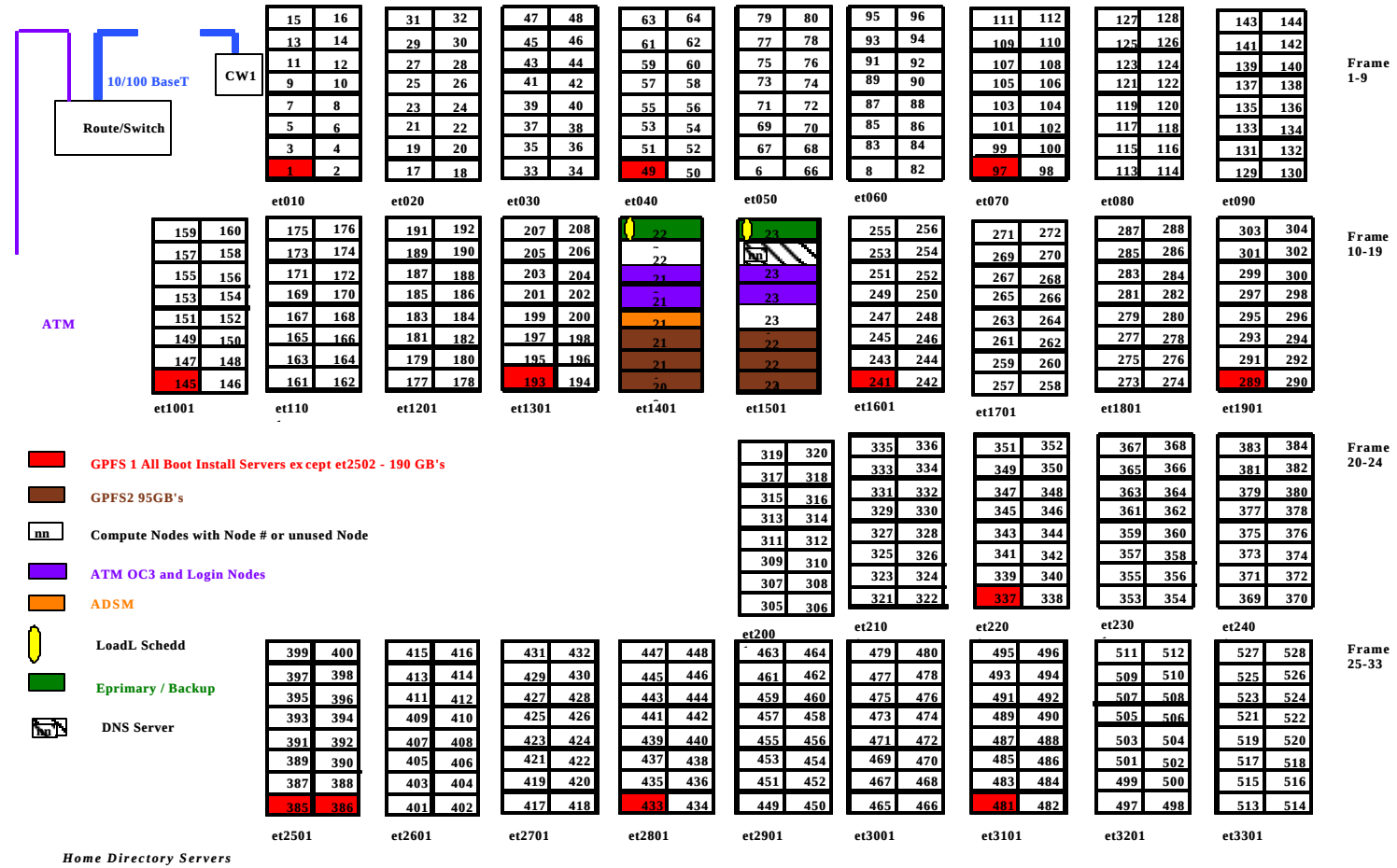
Publications or presentations that result from research which has used resources of the Molecular Science Computing Facility should include the following acknowledgement: This research was performed in part using the Molecular Science Computing Facility (MSCF) in the William R. Wiley Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory. The MSCF is funded by the Office of Biological and Environmental Research in the U.S. Department of Energy. Pacific Northwest is operated by Battelle for the U.S. Department of Energy under Contract DE-AC06-76RLO 1830.

* Current Address is: Federal Express, Colorado Springs, Colorado
* *Contact Michael R. Valley (mike.valley@pnl.gov) for questions or comments
***Currently in Portland, Oregon

July 1999
Updated June 13, 2000

Prepared for
The U.S. Department of Energy
under Contract DE-AC06-76RLO 1830
Pacific Northwest National Laboratory
Richland, Washington 99352

Disclaimer



NWMPP1 Diagram

By Michael Valley
May 25, 2000

Contents

I. Introduction - Purpose / Objectives

The first and foremost task an administrator has to perform on an RS/6000 SP is to set it up. Setting up an RS/6000 SP system requires a fair amount of planning. This guide explains the processes involved in installing and configuring the RS/6000 SP system from the software perspective. Hardware installation is planned with the help of your IBM hardware engineers. They will be able to assist you in developing a satisfactory design.

Concept of SP Installation

The implementation of an RS/6000 SP system involves two types of installations: the hardware and the software. For the hardware installation, an IBM hardware engineer will perform the initial set up. However, you need to understand how the various hardware components interact with one another in order to understand how the software works. The software installation includes two main components: the control workstation and the processor nodes. The installation of the control workstation plays the primary role. Only when this is in place can the rest of the system (the processor nodes) be installed.

In order to achieve a successful implementation, you will need to properly design your final setup. Aspects that you will have to consider are:

- Network design
- The physical equipment and its operational software
- Operational environments
- System partitions
- Migration and coexistence on existing systems
- Security and authentication
- Defining user accounts
- Backup procedures

If you are new to the RS/6000 SP, we recommend that you read IBM RS/6000 SP: Planning, Volume 1, Hardware and Physical Environment, GA22-7280 and IBM RS/6000 SP: Planning, Volume 2, Control Workstation and Software Environment, GA22-7281 for complete information regarding system planning.

Role of Control Workstation and Boot/Install Servers

The control workstation plays the main role in an RS/6000 SP installation. It is from this system that the rest of the nodes are installed. The control workstation is an RS/6000 system running AIX. In order to perform its role as a control workstation, it has to be loaded with the IBM Parallel System Support Programs for AIX (PSSP) software package. With this software installed, it serves as a point of control for installing, managing, monitoring and maintaining the frames and nodes. In addition to these functions, the control workstation is normally set up as an authentication server. After installing at least one node, it can configure one or more of the nodes to become boot/install servers. These nodes, serving as boot/install servers, can be used to boot up and install other nodes, offloading from the control workstation this time-consuming and process-hungry task.

NIM Concept

The core of the SP node installation centers around the Network Installation Management (NIM) in AIX. With NIM, you can manage standalone, diskless, and dataless systems. In a broad sense, an SP node can be considered a set of standalone systems: each node has the capability of booting up on its own because that the RS/6000 SP system is based on a share-nothing architecture. Each node is basically a standalone system configured into the SP frame. Working together with the System Data Repository (SDR), NIM allows you to install a group of nodes with a common configuration or individually customize to each node's requirements. This helps to keep administrative jobs simpler by having a standard rootvg image and a standard procedure for installing a node. The system volume groups are consistent across nodes (at least when they are newly installed). You also have the option to customize an installation to cater to the specific needs of a given node if it differs from the rest.

As NIM installations utilize the network, the number of machines you can install simultaneously depends on the throughput of your network (namely, Administrative Ethernet). Other factors that can restrict the number of installations at a time are the disk access throughput of the installation servers, and the processor type of your servers.

The control workstation and boot/install servers are considered NIM masters. They provide resources (like files, programs and booting capability) to nodes. Nodes are considered NIM clients, as they are dependent on the masters for services. Boot/install servers are both masters and clients since they serve other nodes but are in turn dependent on the control workstation for resources. Not considering NIM masters outside of the SP complex, the control workstation is configured only as a NIM master. The master and clients make up a NIM environment. Each NIM environment can have only one NIM master. In the case of the control workstation and the boot/install server, only the control workstation is considered the master. In another NIM environment consisting of the boot/install server and its client nodes, only the boot/install server is the master.

A NIM master makes use of the Network File System (NFS) utility to share resources with clients. As such, all resources required by clients must be local file systems on the master.
Administrators might be tempted to NFS-mount file systems from another machine to the control workstation as /spdata/sys1/install/images to store node images, and as /spdata/sys1/install/<name>/lppsource to store AIX filesets due to disk space constraints. This is strictly not allowed, as NIM will not be able to NFS export these file systems.

Installation Procedure

The installation of the RS/6000 SP system involves the following steps:

1. Install and configure AIX on the control workstation.
   Note: the control workstation must be at the highest AIX and PSSP level for the whole system.
2. Setting up the SDR information for the frames, nodes and switches.
3. Installation of the nodes.

The rest of this document describes in detail the procedure to set up the RS/6000 SP system. For further details, refer to PSSP: Installation and Migration Guide, GA22-7347.

For a list of related books and information about accessing online information, see the bibliography (References) in the back of the document. This document applies to PSSP Version 3 Release 1.

- To find out what version of PSSP is running on your control workstation (node 0), enter the following: splst_versions -t -n0. In response, the system displays something similar to: 0 PSSP-3.1 if the response indicates PSSP-3.1, this document applies to the version of PSSP that is running on your system.
- To find out what version of PSSP is running on the nodes of your system, enter the following from your control workstation: splst_versions -t -G. In response, the system displays something similar to:

  1 PSSP-3.1
  2 PSSP-3.1
  7 PSSP-2.4
  8 PSSP-2.2

If the response indicates PSSP-3.1, this document applies to the version of PSSP that is running on those nodes.

Who Should Use This Document

This document is intended for system administrators responsible for installing, configuring, and maintaining the RS/6000 SP system. It assumes the administrators have a working knowledge of AIX or UNIX and experience with network systems. The System Administrators Guild of USENIX (SAGE), has developed a classification for skills required for system administrators. Administrators of RS/6000 SP systems are expected to have level II skills (Junior System Administrator) or greater depending on the complexity of your site and system. See Appendix E, "SAGE Job Descriptions" on page 241 of the "Installation and Migration Guide" for PSSP 3.1 for more information on these job skills.

This guide applies to all RS/6000 SP's running AIX V4.3.2 and PSSP 3.1.

How This Document is Organized

I. Planning and Preparation - PSSP 3.1;AIX V4.3.2 or Greater
   - A. Installing and Configuring a New RS/6000 SP System
   - B. Network Planning
   - C. Configuration Files - PRESERVE CURRENT CONFIGURATION
   - D. Filesystem Layout
   - E. Update /systems/lpp from "CD-ROM or tape" to ensure latest base file sets
   - F. Update /systems/ptf from "fixdist" to ensure you have the latest PTFs

II. Prepare the Control Workstation
   - A. BOS Install from AIX V4.3.2 for 5765-C34 CD

III. Setup Nodes to be Installed
   - A. Customize the Nodes
   - B. Setup the Switch
   - C. Power on and Install the Nodes

IV. Customization

V. References - Documents used

VI. Appendix - Scripts used
- A. Scripts In This Appendix Are Protected
- B. The “genders” file
- C. The customize and customize.local scripts
- D. Install.afs; Unconfigure DCE; Make DCE Scripts

II. Planning and Preparation - PSSP 3.1;AIX V4.3.2 or Greater

A. Installing and Configuring a New RS/6000 SP System

1. This document is to be used to do a complete re-install, not for migration.
    a) Please note that *all commands are in* **bold.** Options and other text for emphasis are *italicized and underscored*. There are a number of dependencies that would be unique to Pacific Northwest National Laboratory. Please review the appendix to familiarize yourself with some of these dependencies. This document assumes you have an external nfs server which has all of the base filesets plus ptfs of all code you intend to install. We call our external filesystem *"/systems"* and currently (04/13/99) it contains the following directory structure: */systems/lpp/<subdir>*

        (where <subdir>="one of the names below")
        adsm
        aix432
        aix433
        atm
        cxx364
        dce21
        dce22
        doc432
        gpfs
        graphics
        hippi
        info43
        loadl21
        maui
        netscape
        nwtest
        perf
        pessl
        pessl31
        piofs
        ppe23
        ppe24
        ssp31
        ssp311
        sws
        sysback
        vac44
        vacpp40
        vast
        xlf610
        xlhpf11
        xlhpf13
        xlhpf14

Also has */systems/ptf/<subdir>*
aix43
aix433
ssp31
ssp311

All *ptfs* associated with AIX (dce, C, etc) are in *aix43*
All *ptfs* associated with PSSP (poe, gpfs, essl, pessl, loadleveler, etc) are in *ssp31*
Also has */systems/bin/*
This directory contains *"becoming"*, *"customize"*, *"customize.local"* and other important scripts for everything to install. See the *"appendix"* for information on some of these scripts.

2. This document is intended for system administrators:
   a) Responsible for installing, configuring, and maintaining the RS/6000 SP system.
   b) It assumes the administrators have a working knowledge of AIX and experience with networked systems.
   c) If you decide to use this procedure, Pacific Northwest National Laboratories assumes NO responsibility for any problems which may be incurred as a result of using this procedure. It is completely your responsibility.
   e) Make sure your configuration worksheets are filled out. Reference IBM RS/6000 SP Planning Volume 2 - GA22-7281-03
   f) Under some situations you will use Perspectives
   g) The URL's in this document may not be up to date.
   http://www.rs6000.ibm.com/resource/aix_resource/sp_books/index.html
3. *Before beginning, consider the following carefully*
   a) Make sure you have *a complete backup* of your current system using whatever methodology you are comfortable with.
   b) Make sure you have */systems* configured correctly with all the necessary code.
   c) Make sure you have the following scripts in */systems/bin* samples are in the *appendix*:
      (1) becoming
      (2) customize
      (3) customize.local
      (4) install.afs
      (5) mkdce22.expect
      (6) mkdceclient.22
      (7) uncfgdce.22
      (8) uncfgdce22.expect
   d) Go through *"customize.local"* with a fine toothcomb and make sure you've covered all the bases.
   e) The "Control Workstation" (CWS) must be completely ready before you can successfully run *"customize.local"*. For example, have you installed LoadLeveler on the CWS? Have you installed the "C" compiler on the CWS?

B. Network Planning

1. *Get the following information for every interface on the system*

   IP Address
   Netmask
   Gateway Address
   Emulated LAN Name - SPNET2
   ATM
   Ethernet
   Switch
   Ifconfig -a > ifconfig.txt ; dsh -va "ifconfig -a" >> ifconfig.txt

C. Configuration Files - PRESERVE CURRENT CONFIGURATION

1. *Save current configuration files into /systems/cfg/. The configuration files are different for the login nodes and for the Control Workstation (i.e. /etc/passwd, /etc/security/…), need copies of differences.*

   LoadL_admin
   LoadL_config
   LoadL_config.local
   LoadL_config.local.batch
   csh.cshrc
   csh.login
   dceunixd.sh
   environment
   licenses-C++,VAST, xlf
   mkuser.default
   mkuser.sys
   ntp.conf
   passwd
   profile
   hosts
   resolv.conf
   netsvc.conf
   netstat -r > netstat.routes
   netstat -i > netstat.interfaces
   no -a > no.out
   /etc/security/user
   /etc/security/passwd
   /u/postman/sysinfo.file
   /etc/security/limits
   roots crontab
   /etc/inittab
   /etc/rc.local
   /wcoll (All Working Collectives)

/usr/local/*
lsfs > /systems/cfg/lsfs.<system_name>
lslpp -i > /systems/cfg/lpps.<system_name>
cd / ; tar -cf <nfs>/junk.tar ./qbank ./loadl ./etc ./var ./usr/local ./wcoll
rsh <node> 'cd / ; tar -cf <nfs>/junk.node.tar ./qbank . /loadl ./etc ./var ./usr/local ./wcoll'

D. Filesystem Layout

1. *Take a snapshot of the current system layout for the following (nwmpp1 03/27/2000 example)*

a) CWS

```
Filesystem    1024-blocks      Free %Used    Iused %Iused Mounted on
/dev/hd4       49152      10268   80%     2774 12% /
/dev/hd2       1867776     193444   90%   52759 12%/usr
/dev/hd9var    81920      46252   44%     8308 41% /var
/dev/hd3       294912     121356   59%      810  2%/tmp
/dev/hd1       16384      15760    4%      292 8%  /home
/dev/scratchlv 8847360 2257248 75% 4368 1%        /scratch
/dev/lv00    8011776    2586860    68%    12611 1%  /spdata
/dev/lv01    16384 15804  4%   31  1%              /var/adm/csd
/dev/lv02    327680     104440    69%    13338     3% /usr/local
/dev/lv03    65536      59264   10%        89     1% /tftpboot
/dev/lv04    278528     259696     7%       622     1% /qbank
/dev/lv05    589824     456664    23%    51219    35% /usr/vice/cache
/dev/lv06    1753088     297004   84%     4850     2% /loadl
AFS           72000000  72000000    0%        0     0% /afs
neo:/u1       58589184  30884764   48%   453397     4% /u1
morpheus:/u2 61079552  20259096   67%   297266     2% /u2
prophet:/systems 33554432  11318964   67% 73202 1% /systems
prophet:/images 16646144   9725916   42%  123   1% /images
```

b) Login Node (03/27/2000 example)

```
/dev/hd4       65536      19460   71%     2341     8% /
/dev/hd2       1875968      14544  100%    41292     9% /usr
/dev/hd9var    65536      28168   58%     1943    12% /var
/dev/hd3       319488     286904   11%     1196     2% /tmp
/dev/hd1       8192       7828    5%       284    14% /home
/dev/lv01      589824  146792 76%  51219    35% /usr/vice/cache
/dev/scratchlv  4194304   2766404   35%   357      1% /scratch
/dev/lv02       8192       7320   11%    31      2% /var/adm/csd
/dev/lv03       122880   66736   46%  1814      6% /usr/local
/dev/lv04       262144     249892    5%       70     1% /loadl
AFS              72000000  72000000    0%        0     0% /afs
neo:/u1          58589184  30884760   48%   453398     4% /u1
morpheus:/u2     61079552  20259064   67%   297267     2% /u2
sw1413:/systems  33554432 11318964 67%  73202     1% /systems
/dev/gpfs2     105578496  78058944   27%    14760    15% /gpfs2
/dev/gpfs1     211132416 159828224   25%     5499     5% /gpfs1
prophet:/systems  33554432  11318964 67%  73202   1% /systems
```

c) Compute Node (03/27/2000 example)

```
/dev/hd4        65536      52772   20%     2040     7% /
/dev/hd2        860160      47912   95%    18472     9% /usr
/dev/hd9var     65536      41172   38%      485     3% /var
```

```
/dev/hd3       32768     29460   11%      182     3% /tmp
/dev/hd1       8192      7828    5%       292    15% /home
/dev/scratchlv 6062080   5562288   9%    325     1% /scratch
/dev/lv01      8192      7224   12%     32     2% /var/adm/csd
/dev/lv02      65536     58800   11%   260     2% /usr/local
/dev/lv03      262144    247120    6%   66     1% /loadl
```

2. *Plan how you want your disk to be used for example*
   a) Do you want to Stripe (/usr/local/sbin/stripe.ksh)
   b) /scratch ("big file support")
   c) Swap
   d) Special use filesystems - /usr/local; /usr/remote

3. ****Make Sure To Check Any README FIRST Documents****

4. *Kill the dceunixd.sh process, and run /systems/bin/rmdce22.expect to remove dce configuration.*

E. *Update /systems/lpp from "cdrom or tape" to ensure latest base file sets in:*
/systems/lpp/adsm *(server is on* `prophet:/systems/lpp/adsm`)
/systems/lpp/aix432
/systems/lpp/aix433
/systems/lpp/cxx364
/systems/lpp/dce22
/systems/lpp/gpfs
/systems/lpp/info43
/systems/lpp/loadl21
/systems/lpp/maui *(CWS ONLY)*
/systems/lpp/perf
/systems/lpp/pessl31 *(Do NOT install on CWS)*
/systems/lpp/ppe24 *(Do NOT install on CWS)*
/systems/lpp/ssp31
/systems/lpp/ssp311
/systems/lpp/sysback
/systems/lpp/vac44 *(Install Everywhere on NWTEST only on Login Nodes and CWS for mpp1)*
/systems/lpp/vacpp40 *(Install Everywhere on NWTEST only on Login Nodes and CWS for mpp1)*
/systems/lpp/vast *(Do NOT install on CWS)*
/systems/lpp/xlf610 *(or latest)*

F. *Update /systems/ptf from "fixdist" to ensure you have the latest ptfs*
aix43 - Any ptfs associated with AIX i.e. C++, DCE, etc.
aix433
ssp31 - contain essl, pessl, xlhpf, gpfs ppe.poe, loadleveler and ssp ptfs
ssp311

III. Prepare the Control Workstation

A. BOS Install from AIX V4.3.2 for 5765-C34 CD

1. ******Make Sure To Check Any "README FIRST" Documents******

2. *Perform Complete Overwrite install*
   http://www.rs6000.ibm.com/doc_link/en_US/a_doc_lib/aixgen/wbinfnav/ListOBooks.htm
   AIX V4.3.2 Installation Guides

3. *Change root password*

4. *Change system time/zone*

5. *Change swap space - 2 x memory*
   Reference IBM Performance and Tuning Guide
   a) The install process creates paging space equal to two times memory (2X) for systems with less than 64MB of RAM. For systems with 64MB of RAM or more, but not greater than 256MB, the paging space is RAM size + 16MB. The following equation is used for systems with more than 256MB of RAM: Page Space = 512 + (RAM - 256) * 1.25. RAM can be found with: **'let A=$(bootinfo -r)/1024 ; echo $A'**. LPSize can be found with: **'let B=$(lsvg rootvg |grep "PP SIZE:" |cut -f3 -d: |tr -d "a-z ()") ; echo $B'**. Number of hd6 LP's can be found with:**'let C=$(lsps -s |tail -1 |awk '{print $1}' |tr -d MB)/$B ; echo $C'.** Additional LP's= ((RAM - 256)*1.25 + RAM) / LPSize - [number of hd6 LPs]
   b) NWTEST (512 MB of memory and 8 MB LP's with 64 MB of swap already defined ((512 - 256) * 1.25 + 512) / 8 - 8 = 96 **extendlv hd6 96 ; swapon /dev/hd6**

6. *smitty install_selectable_all*
   a) /systems/lpp/aix<level> (source)
   b) press **F4** to list program products
   The following is the minimal list of AIX file sets required:

bos	xlC.rte.*
bos.diag.*	X11.apps.*
X11.base.*	X11.compat.*
X11.Dt.*	X11.fnt.*
X11.loc.En_US.*	X11.motif.*
X11.msg.En_US.*	X11.vsm.*
bos.mp.*	bos.net.*
bos.powermgt.*	bos.sysmgt.*
bos.terminfo.*	bos.up.*
bos.64bit	devices.*
perfagent	bos.atm.*

- After the Base Operating System (BOS) is installed, you may want to install optional software or service updates. This section discusses software installation and applying service updates, but does not discuss committing service updates after installation. For information on committing, rejecting, and removing software after installation, refer to "Maintaining Optional Software" and "Optional Software Installation and Update Concepts".
- For information on cleaning up after an interrupted software installation, refer to "Cleaning Up Optional Software and Service Updates".

Optionally installed software includes the following:

- Optional Software Products. An optional software product is software that is not automatically installed on your system when you install BOS.
- Service Updates. A service update is software that corrects a defect in or adds new function to the BOS or an optional software product.
- Service updates are organized by filesets. Filesets are sets of files that are part of the same optional software product. Some products are not organized as fileset updates. Such products can only be updated by installing a newer version.
- For more information on updating software, see "Optional Software Installation and Update Concepts".

Installing and Running Hardware Diagnostics

- If your system is not equipped with a CD-ROM drive, install the hardware diagnostics bundle from your installation media to enable concurrent hardware diagnostics.
- If your system is equipped with a CD-ROM drive, use the diag command to run concurrent diagnostics from the diagnostics CD-ROM. Run diagnostics in the standalone mode by booting from the diagnostics CD-ROM.

7. *Setup the primary ethernet interface*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.6
Step 6: Tune All Control Workstation Network Adapters Page 14, PSSP Installation Guide PSSP 3.1

a) Tune All Control Workstation Network Adapters
   (1) PCI - 256, MCA - 512
   (2) **smitty chdev** or
   (3) **chdev -P -l ent0 -a xmt_que_size=512** (256 or 512)

8. *Configure Network Interfaces - nwmpp1 example, March 28, 2000*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.7
Step 7, PSSP Installation Guide

a) **smitty mktcpip** (Ethernet)
(1) en0 - Example:
   Ensure the hostname is the fully qualified 'hostname'
   HOSTNAME [cw1.nwmpp1.emsl.pnl.gov]
   Internet ADDRESS (dotted decimal) [198.129.130.1]
   Network MASK (dotted decimal) [255.255.255.224]
   Default GATEWAY Address [198.128.88.1]
b) Or use the following command Example:
(1) **/usr/sbin/mktcpip -h 'cw1.nwmpp1.emsl.pnl.gov' -a '198.129.130.1 ' -m'255.255.255.224' -i'en0' -g'198.128.88.1 -t'N/A'**
(2) **shutdown -Fr**

(3) **netstat -v (**Get the atm interface MAC address)
c) **smitty atmle_panel** (ATM)
(1) Add an ATM LE Client
(2) Add an Ethernet ATM LE Client
(3) Info to enter:
MAC Address (from step 5 above)
Auto config *yes*
emulated LAN name *spnet2*
d) Setup other ethernet network interfaces (en1, en2, etc) using
e) **smitty chinet**
(1) Add A Standard Ethernet Network Interface
(2) Info to enter:
Select en1
Enter IP address (Supplied by network admin)
Enter netmask (Supplied by network admin)
f) **smitty route**
(1) Add a static route
Destination type: net
Destination address: default
Default GATEWAY Address: (Supplied by network admin)
g) check default route **(netstat -rn)**
ping default gateway
Set the hostname for the ATM interface
Ensure this is not the same as the default 'hostname'
h) mkdir /systems
i) **smitty mknfsmnt**
Add a file system for mounting
pathname: /systems
pathname of remote directory
host: 198.129.136.5 (03/27/2000 prophet.emsl.pnl.gov)
mount now: both
j) *Make "SURE" you have a clean hosts file with the correct information.*
(1) **cp -p /systems/cfg/etc/hosts.<machine_name> /etc/hosts**
k) **cp -p /systems/cfg/etc/resolv.conf.<machine_name> /etc or vi /etc/resolv.conf**
(1) Add secondary nameserver
(2) nameserver 130.20.84.36
Sample entry in /etc/reolv.conf on *cw1* is:
Search nwmpp1.emsl.pnl.gov emsl.pnl.gov pnl.gov
nameserver 198.128.84.27
nameserver 130.20.20.36
nameserver 130.20.84.36
l) **cp -p /systems/cfg/etc/netsvc.conf /etc/**
m) Ensure hostname is fully qualified and is associated with the primary ethernet (en0) adapter (**smitty mktcpip**)

9. *Verify the CWS Interfaces (Step 8, Page 16 PSSP Installation Guide)*
a) **ping -c 1 198.129.130.1**

10. *Ensure Necessary Daemons Are Running on the CWS*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.9 Step 9, PSSP Installation Guide

a) **lssrc -a**

11. *Remove unnecessary file sets*

a) Non-US English file sets (bos.msg;bos.loc)

12. *Install additional file sets on the CWS*

a) /systems/lpp/perf/install.perf
b) /systems/lpp/dce22/install.dce
c) /systems/lpp/loadl21/install.loadl
d) /systems/lpp/vac44/install.vac
e) /systems/lpp/vacpp40/install.cpp
f) /systems/lpp/cxx364/install.cxx364
g) /systems/lpp/doc432/install.man
h) /systems/lpp/sysback/install.update
i) shutdown -Fr

13. *Do a SYSBACK or MKSYSB of the CWS; Label and Put Away*

14. *Do MKSYSB to Disk or Use " /systems/lpp/ssp31/spimg.usr.3.1.0.0"*

a) If you are going to use the "spimg", *skip steps b-f* below. Continue with *step g*.
b) Create a Filesystem large enough to hold a mksysb, about 500MB
(1) This filesystem needs to be in it's own Volume Group (**smitty mkvg**) call it imagevg. Call the filesystem *images. (***smitty crjfs***)*
c) Create mksysb to the new filesystem called *bos.obj.compute* (smitty mksysb) *Make sure LoadLeveler is NOT part of the mksysb*. Use /etc/exclude.rootvg and specify */usr/lpp/LoadL/* in the file.
d) Copy the mksysb node image to tape
(1) **cd /images**
(2) tar -cvf /dev/rmt0 bos.obj.compute
(3) Label this tape "Node Image for <machine_name>"
e) Umount and rmfs the new filesystem.
(1) **umount /images**
(2) **rmfs /images**
f) **varyoffvg imagevg;exportvg imagevg**
g) Create a mksysb or "sysback" of the cws to tape (**smitty mksysb or smitty sysback**).
h) Label this tape "Bootable backup of <cws_name>
i) Create a mksysb or sysback at the end of each day with a new date

15. *BEGIN SSP INSTALLATION PROCEDURE*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.1 Step 1: Update the root User Path Steps 1-2, Page 12 PSSP Installation Guide

a) **cp -p /systems/cfg/profile.cws /.profile**
b) **check permissions**
c) **/.profile**
d) **cp -p /systems/cfg/etc/environment /etc/**

e) **/usr/lib/instl/inurid -q ; echo $?**
(1) Verify that it returns 0.
(2) If not, reinstall cws. (Arghhhhh!!)

16. *Configure rs232 interfaces*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.5
Step 5, Page 14 PSSP Installation Guide

a) **smitty tty**
(1) Add a tty
(2) select the appropriate tty adapter.
(3) select the appropriate interface
(4) NWTEST
  (a) Command line: **mkdev -c tty -t 'tty' -s 'rs232' -p 'sa0' -w 's1'**
(5) All other options stay at default values.
b) **vi /etc/inetd.conf**
c) Search for bootps and tftp (/bootps /tftp).
d) If either of these are commented, uncomment them.
e) refresh -s inetd

17. *Change CWS system environment*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.10
Step 10, Page 17 PSSP Installation Guide)

a) **smitty system**
(1) Change / Show Characteristics of Operating System
(2) Maximum number of PROCESSES allowed per user[256]
  (a) Command line: **chdev -l sys0 -a maxuproc='256'**
(3) Change / Show Number of Licensed Users
(4) Maximum number of FIXED licenses [65]
  (a) Command line: **chlicense -u '65'**

18. *Workstation tunables*

(Step 11, Page 17 PSSP Installation Guide)

a) SP System Specific Tunable Recommendations "IBM Redbook SG24-5340"
http://www.redbooks.ibm.com/pubs/html/redbooks/
(1) *thewall* - set to at least 25% of real memory. Current value NWmpp1 is 65536.
(2) *sb_max* - This value should be at least twice the size of the largest value for *tcp_sendspace*, tcp_recvspace, udp_sendspace, and udp_recvspace. Current value for the CWS on NWmpp1 is 2097152; for a compute/login node is 4194304
(3) *tcp_sendspace* - Minimum of 65536 unless there are problems with external connections not handling windows bigger than 32768. Never higher than the major network adapter transmit queue limit. To calculate this limit, use:Adapter Queue Size * Smallest Network Adapter MTU. Current value on the CWS for NWmpp1 is 655360; on a compute/login node is 262144
(4) *tcp_recvspace* - Minimum of 65536 (same as tcp_sendspace). Current value on NWmpp1 is 262144.
(5) *udp_sendspace* - Set to 65536 because anything beyond 65536 is ineffective.
(6) *udp_recvspace* - A suggestion for a starting value for udp_recvspace is 10 times the value of udp_sendspace. For large parallel application using UDP, the value may have to be increased. Current value on NWmpp1 is 655360.

(7) *rfc1323* - Always set to on (1).
(8) *tcp_mssdflt* - This tunable is used to set the maximum packet size for communication with remote networks. This is calculated by:
(9) MTU of interface - TCP header size - IP header size -rfc1323 header size = MTU -20 -20-12 = MTU-52
(10) Limiting data to the MTU-52 bytes ensures that, where possible, only full packets will be sent.
  (a) Check */etc/rc.net* on the CWS and make sure the values correspond to the correct machine. On the nodes you need to check */etc/rc.local*.
  Ensure no values for the CWS and "boot install servers" are:
  /usr/sbin/no -o tcp_sendspace=65536
  /usr/sbin/no -o tcp_recvspace=65536
  /usr/sbin/no -o thewall=65536
  /usr/sbin/no -o sb_max=2097152
  /usr/sbin/no -o ipforwarding=1
  /usr/sbin/no -o udp_sendspace=65536
  /usr/sbin/no -o udp_recvspace=655360
  /usr/sbin/no -o tcp_mssdflt=1448
  /usr/sbin/no -o rfc1323=1

19. *Define space for NIM boot images*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.13
Step 12, Page 18 PSSP Installation Guide

a) Create /tftpboot filesystem (~25MB=50,000 512 byte blocks)
(1) **smitty crjfsstd**
(2) **mount automatically = yes**
(3) **mount /tftpboot**

20. *Define space for /spdata*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.13
Steps 13.1-13.4, Pages 19, 20 PSSP Installation Guide

a) Create volume group (nwmpp1 example)
(1) Mirror rootvg (Skip this step if your are not mirroring)
  (a) **smitty extendvg** (include hdisk2,hdisk3 for nwmpp1)
    (1) smitty mkvg (Done for every install)
    VOLUME GROUP name [spvg]
    Physical partition SIZE in megabytes 8
    (PP Size depends on the size of the disk drives, i.e. 4.5GB drives require 8mb pp size; 9GB drives require 16mb pp size)
    PHYSICAL VOLUME names <all_unused>]
    Command line: mkvg -f -y spvg -s 8 hdisk#
b) Create a logical volume (called "splv")
(1) **smitty mklv**
  (a) select # of physical partitions to use almost (minus a couple of physical partitions) the entire disk. To determine the number of pp's available do: lsvg spvg and look at the right most column.
  (b) Command line: **mklv -y splv -x 530 spvg 530** (where # of pp's from lsvg spvg =530)

c) create a file system

(1) **smitty crfs**

(a) create a file system over a "previously defined logical volume"

(b) Add a Large File Enabled Journaled File System

LOGICAL VOLUME name splv
MOUNT POINT [/spdata]
Mount AUTOMATICALLY at system restart? Yes

(2) Command Line: **crfs -v jfs -d splv -g rootvg -m /spdata -A yes -p rw -t no -a frag=512 -a npbi=512 -a ag=8**

(3) **mount /spdata**

21. *Change system filesystem sizes to useable sizes. Please NOTE - In PSSP 3.1 Customizations*

a) The /(root) filesystem *has been increased to 16MB*.

b) The /var filesystem *should be increased to 50MB***.**

c) Default number of licensed users *was set to >64*

d) Default maximum number of processes allowed per user *was set to 256*

e) Restrictions/Considerations: You should *clean out /tmp* after the image is installed, so there is maximum working space.

f) AIX X-Windows has been removed (to reduce the image size) from the spimg in this release.

g) After you have installed your nodes, you should *increase the dump space* on each node to *at least 32 MB***.** This is the minimum dump space required for the minimal image pre-installed on each SP node and supplied in the "spimg" installp image. If you are using a larger image, more dump space may be required. You may cause the dump space on a node to be increased during installation by specifying the extendlv command in the script.cust script. If script.cust is run multiple times (i.e. customization of nodes) then all statements will also be run multiple times. Refer to the sample script.cust in /usr/lpp/ssp/samples

h) NWTEST filesizes should be about (df -k snapshot):

/dev/hd4	20480	10020	52%	1789	18% /
/dev/hd2	851968	10952	99%	28847	14% /usr
/dev/hd9var	24576	19204	22%	345	6% /var
/dev/tftpbootlv	25536	6764	90%	57	1% /tftpboot
/dev/lv00	229376	87416	62%	9410	5% /usr/local
/dev/lv02	434176	191628	56%	861	1% /loadl
/dev/hd3	24576	22996	7%	68	2% /tmp
/dev/splv	4390912	482240	90%	20137	2% /spdata

i) NWMPP1 filesizes should be about:

/dev/hd4	40960	25140	39%	2402	12% /
/dev/hd2	827392	22184	98%	24787	12% /usr
/dev/hd9var	98304	23248	77%	2226	10% /var
/dev/hd3	40960	34376	17%	320	4% /tmp
/dev/hd1	8192	7828	5%	50	3% /home
/dev/tftpbootlv	65536	6764	90%	57	1% /tftpboot
/dev/lv00	229376	87416	62%	9410	5% /usr/local
/dev/lv01	81920	61304	26%	17	1% /var/adm/cacct
/dev/lv02	434176	191628	56%	861	1% /loadl

/dev/splv	4390912	482240	90%	20137	2% /spdata
/dev/banklv	204800	175612	15%	191	1% /qbank

22. *Create the required spdata directories (make sure directories have permissions of* rwxr-sr-x*)*
    a) **mkdir -p /spdata/sys1/install/aix432/lppsource**
    b) **mkdir /spdata/sys1/install/images**
    c) **mkdir -p /spdata/sys1/install/pssplpp/PSSP-3.1**
    d) **mkdir /spdata/ sys1/install/pssp**

23. *Copy the lpp's into the lppsource directory*
    a) Required file sets are:

bos
bos.diag.*
bos.mp.*
bos.net.*
bos.powermgt.*
bos.sysmgt.*
bos.terminfo.*
bos.up.*
bos.64bit
devices.*
perfagent
xlC.rte.*
X11.apps.*
X11.base.*
X11.compat.*
X11.Dt.*
X11.fnt.*
X11.loc.*
X11.motif.*
X11.msg.*
X11.vsm.*

(http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.15 Step 15, Page 22 see table PSSP Installation Guide)

Use these commands as the parameter list is too long for a normal copy:

    a) **find /systems/lpp/aix432 -type f -exec cp -p {} /spdata/sys1/install/aix432/lppsource \;**
    b) **find /systems/ptf/aix43 -type f -exec cp -p {} /spdata/sys1/install/aix432/lppsource \;**
    c) **cp -p /systems/lpp/perf/* /spdata/sys1/install/aix432/lppsource**
    d) **cp -p /systems/lpp/ssp31/* /spdata/sys1/install/pssplpp/PSSP-3.1**
    e) **cp -p /systems/lpp/gpfs/* /spdata/sys1/install/pssplpp/PSSP-3.1**
    f) **cp -p /systems/ptf/ssp31/* /spdata/sys1/install/pssplpp/PSSP-3.1**
    g) **inutoc /spdata/sys1/install/aix432/lppsource**

24. *INSTALL PSSP and setup CWS*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.3.1.1 Step 16, Page 25 PSSP Installation Guide

    a) Make sure the following images are in the /systems/lpp/ssp31 directory:

- *ssp.vsdgui* Contains the SP Virtual Shared Disk Perspective
- *ssp.loc* Contains the non-english locale information for the SP Perspectives Launch Pad, SP Hardware Perspective, and SP Event Perspective
- *ssp.msg* Contains the non-english messages for the SP Perspectives Launch Pad, SP Hardware Perspective, and SP Event Perspective
- *ssp.help* Contains the non-english online help for the SP Perspectives
- *ssp.top.loc* Contains the non-english locale info for the SP System Partitioning Aid
- *ssp.top.msg* Contains the non-english messages for the SP System Partitioning Aid

- *ssp.ptpegui.loc* Contains the non-english locale info for the SP Performance Monitor Perspective
- *ssp.ptpegui.msg* Contains the non-english messages for the SP Performance Monitor Perspective
- *ssp.vsdgui.loc* Contains the non-english locale information the SP Virtual Shared Disk Perspective
- *ssp.vsdgui.msg* Contains the non-english messages for the SP Virtual Shared Disk Perspective

b) **mv /spdata/sys1/install/pssplpp/PSSP-3.1/ssp.usr.3.1.0.0 /spdata/sys1/install/pssplpp/PSSP-3.1/pssp.installp**

c) **mv /spdata/sys1/install/pssplpp/PSSP-3.1/rsct.basic.usr.1.1.0.0 /spdata/sys1/install/pssplpp/PSSP-3.1/ rsct.basic**

d) **mv /spdata/sys1/install/pssplpp/PSSP-3.1/rsct.usr.clients.usr.1.1.0.0 /spdata/sys1/install/pssplpp/PSSP-3.1/ rsct.clients**

e) **inutoc /spdata/sys1/install/pssplpp/PSSP-3.1**

25. *Copy the SP Image ("spimg") or bos.obj.compute to /spdata/sys1/install/images*

a) **cd /spdata/sys1/install/images**

b) **smitty install**

(1) /systems/lpp/ssp31/spimg.usr.3.1.0.0 .

c) or ***tar -xvf /dev/rmt0 bos.obj.compute*** from the tape labeled "node image for <machine_name>"

26. *Install ssp on the cws*

a) **smitty install_latest**

(1) Install from */spdata/sys1/install/pssplpp/PSSP-3.1*

(2) Take defaults

27. *Install GPFS on CWS*

a) **cd /systems/lpp/gpfs;./install.gpfs**

28. *Apply Latest PTF's to AIX43*

a) **cd /spdata/sys1/install/aix432/lppsource** and run smitty update_all

b) Reboot the CWS when the update is finished

29. *Do a SysBack (**smitty sysback**)*

a) Include rootvg and spvg in the backup

30. *Setup Authentication Services*

http://ppdbooks.pok.ibm.com:80/cgi-in/bookmgr/bookmgr.cmd/books/sspim310/2.3.7 Step 19, Page 30 PSSP Installation Guide

a) Prior to doing the steps, you must decide what type of authentication is required: See *SP Planning, Vol. 2, Page 142-144, Planning for Security.*

(1) SP, AFS, or another MIT Kerberos Version 4 or 5 (DCE)

(a) For Kerberos 5 Authentication with DCE you need to:

(i) Decide which cell to configure authentication in.

(ii) Be authorized to install and configure DCE.

(iii) Plan the use of AIX Remote Commands.

(iv) Install and configure DCE on the CWS.
(v) Enable the Kerberos 5 authentication method.

(2) Configuring DCE
(a) Make sure to install DCE using */systems/lpp/dce22/install.dce*
(b) First unconfigure DCE (assuming the SP had been configured before) using */systems/bin/rmdce22.expect*
(i) You must check that cell_admin's password is correct in /systems/bin/uncfgdce22.expect.
(ii) The **cp -p /systems/bin/rmdce22.expect /tmp**
(iii)**/tmp/rmdce22.expect**
(c) Then configure by running */systems/bin/mkdce22.expect*
(i) You must check that cell_admin's password is correct in /systems/bin/mkdce22.expect.
(ii) Check to insure the variables have the proper values in /systems/bin/mkdceclient.22
(iii)**cp -p /systems/bin/mkdce22.expect /tmp**
(iv) **/tmp/mkdce22.expect**

(3) **setup_authent**
(a) Decide and *"RECORD"* the master password
(b) Principal name: *root*
(c) Instance: *admin*
(d) Instance: *admin* <Not found>, Create [y]?
(e) <Enter>
(f) *New Password: <rootPassword>*
(g) Verify, please re-enter New Password: *rootPassword Expiration Date [2037-12-31]*
(h) Max ticket lifetime (*5 minutes) [255] ? 255
(i) <Enter>
(j) <Enter>
(k) Kerberos Initialization for "root.admin" Password: *rootPassword*

(4) **# k4list**
(a) Ticket file: /tmp/tkt0
(b) Principal: root.admin@NWMPP1.EMSL.PNL.GOV
(c) Expires
(d) Principal <date/time>
(e) Krbtgt.NWMPP1.EMSL.PNL.GOV (or NWTEST, ECS1)

31. *Complete SSP Install on the cws*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.3.8
Step 20-22, Page 39-40 PSSP Installation Guide

a) **install_cw**
b) **perspectives &** (go ahead and run perspectives to test the graphics, etc.)
c) **SDR_test**
d) **spmon_itest**

32. *Enter Site Environment, Frame, Node, and Switch Information*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.1
Step 23, Page 40 PSSP Installation Guid

a) **smitty enter_data**

(1) Default Network Install Image [bos.obj.compute]

Remove Install Image after Installs	true
NTP Installation	timemaster
NTP Server Hostname(s) [nighthawk.emsl.pnl.gov,skyhawk.emsl.pnl.gov, mohawk.emsl.pnl.gov>	
NTP Version	3
Automounter Configuration	false
Print Management Configuration	false
Print system secure mode login name	[""]
User Administration Interface	false
Password File Server Hostname	[cw3]
Password File	[/etc/passwd]
Home Directory Server Hostname	[cw3]
Home Directory Path	[/home/cw3]
File Collection Management	true
File Collection daemon uid	[102]
File Collection daemon port	[8431]
SP Accounting Enabled	false
SP Accounting Active Node Threshold	[80]
SP Exclusive Use Accounting Enabled	false
Accounting Master	[0]
Control Workstation LPP Source Name	[aix432]

Make sure the items above reflect the values below
remove images after installing: true
image name: bos.obj.compute
automounter configuration: false
set lppsource to aix432.

33. *Enter Frame information*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.2
Step 24, Page 41 PSSP Installation Guide

a) **smitty enter_data**

(1) SP Frame Information

(2) set appropriate values and reinitialize the SDR.

b) Command line: **spframe -r yes 1 1 /dev/tty0**

c) Command line: **spframe -r yes 1 33 /dev/tty0**

34. ***spmon_ctest*** *(verify system monitor information)*

35. ***spmon -d*** *(verify frame information)*

36. *Update the state of frame supervisor code -* ***smitty supervisor***
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.5
Step 27, Page 45 PSSP Installation Guide

a) Check for supervisors that require further action.
(1) Or run the command **spsvrmgr -G -r status all**

37. ***smitty node_data*** *(Enter required node information.)*

a) sp ethernet info
start frame 1
start slot 1
node count (appropriate number - 8, 16, 512)
Starting Node's en0 Hostname or IP Address
Default Route Hostname or IP Address
skip ip addresses: no.
b) **splstdata -n** (verify node objects have been created)

38. *smitty node_data*

a) Create or restoring /etc/bootptab.info
(1) Use the following format (example):
NODE# MAC_ADDRESS
1 08005ABAB177
2 08005ABAAEAB
b) get hardware ethernet addresses
start frame 1
start slot 1
node count (8 for nwtest, 512 for nwmpp1)
(1) Command line: **sphrdwrad 1 1 8**

39. *Verify ethernet addresses (**splstdata -b**)*
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.8
Step 30, Page 50 PSSP Installation Guide

40. *Configure Initial Hostnames* ***(smitty node_data)***

a) Do this step if you do not want the default hostname to match the en0 adapter name or
b) You are using short host names.

41. *smitty node_data*

a) additional Adapter Information (fill in information as required for the switch)
start frame 1
start slot 1
node count (appropriate number, 8 for nwtest, 512 for nwmpp1)
adapter name (css0)
Starting Node's IP Address or Hostname

skip ip: no
enable ARP: yes
Use switch node numbers: no

42. ***splstdata -a*** *(to verify that switch names have been set)*

43. *Select Authorization Methods*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.11
Step 33, Page 55 PSSP Installation Guide

a) **smitty spauth_config**
Select Authorization Methods
Select partition name
Select "k4" and "std" authentication methods Security
Considerations

b) If you wish to use Kerberos 5 authentication, you must perform the following steps prior to enabling the authentication method (Step 34 in the PSSP Installation & Migration Guide): *Install and configure DCE authentication on the CWS and all nodes in the partition.*

(1) Create a .k5login file containing DCE principals for the CWS and all nodes in the partition.

(2) To enable root rcmd access, *set the KRB5CCNAME* environment variable to point to the DCE credentials file. For example: *export KRB5CCNAME=FILE:/opt/dcelocal/var/security/creds/dcecred_ffffffff*

Command line (NWtest): /usr/lpp/ssp/bin/spsetauth -d -p cw3 k4 std

44. *Enable Authentication Methods*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.12
Step 34, Page 56 PSSP Installation Guide

a) smitty spauth_config
Enable Authentication Methods
Enable on CWS only - yes
Force change on nodes - no
Select partition name
Select "k4" and "std" authentication methods

b) Command line: **chauthent -k4 -std**
Command line: **chauthpar -c -p <cw name> k4 std**

45. *Start System Partition-Sensitive Subsystems*

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.14
Step 36, Page 57 PSSP Installation Guide

a) **syspar_ctrl -A**

46. *Verify that System Partition-Sensitive Subsystems Have Started*

a) **lssrc –a**

IV. Setup Nodes to be Installed

http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.17 Step 39, Page 58 PSSP Installation Guide

47. *Make sure "site environment" (***smitty site_env_dialog***) is set such that:*
    a) install image = bos.obj.compute
    b) lpp source name = aix432
    - Do this step if you want to change the default installation settings for any of the nodes. To find out the default settings of your nodes, use the splstdata command. Be aware that the root password is not set if you are installing from a minimal mksysb. For more information on setting a root password when installing from a minimal mksysb, refer to "*Step 44: Perform Additional Node Customization" in topic 2.5.1. PSSP Installation Guide"*
    - If the boot/install server will be forwarding packets from the control workstation to a client node, the boot/install server is acting as a gateway to the control workstation. Therefore, ipforwarding must be correctly enabled. To turn ipforwarding on, issue: */usr/sbin/no -o ipforwarding=1*
    - You cannot export /usr or any directories below /usr because an NFS export problem will occur**.** If you have exported the /spdata/sys1/install/image directory or any parent directory, you must unexport it using the exportfs -u command before running setup_server. You need to do this because NIM attempts to export /spdata/sys1/install/images/bos.obj.ssp.*, where bos.obj.ssp.* is the install image during setup_server processing. If you do not perform this task, you will receive an error. See the "Diagnosing NIM Problems" chapter in the PSSP: Diagnosis Guide for more information.
    - Everything that is required in PSSP is installed on the nodes automatically, regardless of whether you use your own mksysb or the SP minimal mksysb.
    - This step does the following: Changes the default boot/install information for the node objects in the SDR so that you can indicate a different boot/install server configuration to the RS/6000 SP system. Allows you to specify an alternate disk or disks to use when installing AIX on nodes.
    - The default installation assumes one of the following: You have only one frame and the control workstation is configured to act as the boot/install server. You have multiple frames, the control workstation, and the first node in each frame is configured to act as boot/install server. The default installation assumes your nodes have not been preinstalled.
    - If you want to have them installed with your own install image, you must specify the following: Which nodes you are installing with your own install image. The name of the installation image you are using if you do not want to use the default image.
    - If you want different nodes to be installed by a different boot/install server, you must specify the target nodes and which node will serve as the boot/install server.
    c) You will be using the command *"spchvgobj"***.** This command is new with PSSP 3.1. Reference the man page:

spchvgobj
Purpose
spchvgobj - Changes the contents of a Volume_Group object.

Syntax
spchvgobj -r volume_group [-h pv_list] [-i install_image]
[-p code_version] [-v lppsource_name] [-n boot_server]
[-c 1 | 2 | 3] [-q true | false]
{start_frame start_slot node_count | -l node_list}]

48. *On nwmpp1, identify the boot-install servers then each frame:*

(1) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 -l 1,17,33,49,65,81,97,113,129,145,161,177,193,209,241,257,273,289,305,321,337,353,369,385,401,417,433,449,465,481,497,513 -n0
(2) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 1 2 16 -n 1
(3) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 2 2 16 -n 17
(4) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 3 2 16 -n 33
(5) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 4 2 16 -n 49
(6) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 5 2 16 -n 65
(7) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 6 2 16 -n 81
(8) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 7 2 16 -n 97
(9) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 8 2 16 -n 113
(10) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 9 2 16 -n 129
(11) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 10 2 16 -n 145
(12) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 11 2 16 -n 161
(13) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 12 2 16 -n 177
(14) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 13 2 16 -n 193
(15) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 14 2 16 -n 209
(16) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 16 2 16 -n 241
(17) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 17 2 16 -n 257
(18) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 18 2 16 -n 273
(19) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 19 2 16 -n 289
(20) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 20 2 16 -n 305
(21) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 21 2 16 -n 321
(22) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 22 2 16 -n 337
(23) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 23 2 16 -n 353
(24) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 24 2 16 -n 369
(25) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 25 2 16 -n 385
(26) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 26 2 16 -n 401
(27) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 27 2 16 -n 417
(28) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 28 2 16 -n 433
(29) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 29 2 16 -n 449
(30) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 30 2 16 -n 465
(31) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 31 2 16 -n 481
(32) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 32 2 16 -n 497
(33) spchvgobj -r rootvg -h hdisk0,hdisk1 -v aix432 33 2 16 -n 513
(34) On NWTEST use:
spchvgobj -r rootvg -h hdisk0 -i bos.obj.compute -v aix432 1 1 8
*** *NOTE: node 0 is the default in a single frame system you do not supply -n0 at the end of the above command. It will fail.* ***

49. ***splstdata -b*** *Verify everything is correct*
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.18

50. *Configure the Control Workstation as the Boot/Install Server* - make sure steps 23 and 24 have been done before you continue
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.21
Step 43, Page 63 PSSP Installation Guide

51. *setup_server*
   a) Check that all required nim objects have been created (**lsnim**)
   b) Make sure */etc/bootptab* contains an entry for each node to install from this BIS
   c) Make sure */tftpboot* conatins 5 files for each node to be installed

52. *Verify that the System Management Tools Were Correctly Installed*
   a) **SYSMAN_test**

A. *Setup the Switch*
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.21

53. *smitty annotator*
* Input Topology File Name [ ]
* Output Topology File Name [ ]
* Save Output File to SDR [yes]
   1. Input Topology File Name [/etc/SP/expected.top.1nsb.0isb.0](NWTEST)
   2. Input Topology File Name [/etc/SP/expected.top.32nsb.16isb.0.prh.anon] (NWMPP1)
   3. Output Topology File Name [expected.top.annotated]
   4. Save Output File to SDR [yes]

54. *Verify the Switch Primary and Primary Backup Nodes*
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.5.4
Step 47, Page 69 PSSP Installation Guide
   a) Eprimary et1405 -backup et1505 (nwmpp1)
   b) Eprimary t01n01 -backup t01n09 (nwtest)

55. *Set the Switch Clock Source for All Switches*
   a) **Eclock -f /etc/SP/Eclock.top.1nsb.0isb.0** (NWTEST)
   b) **Eclock -f /etc/SP/Eclock.top.32nsb.16isb.0.prh.anon** (NWMPP1)
   c) **splstdata -s** (verify the switch information)

B. Power on and Install the Nodes
*******NOTE: If you are reinstalling the nodes with the "spimag" after a failed installation, return to Section B, step 50. *******

56. *perspectives &*
   a) If you are working remotely, set your display (**setenv DISPLAY <hostname>:0.0** for csh; **export Display=<hostname>:0.0** for ksh)
   b) Hardware ("for many nodes" on nwmpp1) Perspectives
   c) Click on frame to netboot
   d) Click on "View" , "select all"
   e) Hold down the Ctrl key and click on any node you "DON'T" want to install

f) Repeat steps b, c, d for two frames. Don't do more than two frames.
g) Click on "Actions", netboot
h) **spled&**
i) **s1term frame node**

57. *SYSMAN_test*

58. ***Eclock -d*** *(on NWTEST this takes about 1.17 minutes)*

59. ***SDRChangeAttrValues switch_responds isolated=0*** *(this makes sure nodes are unfenced in the SDR)*

60. ***Estart*** *(on NWTEST this takes about 1.11 minutes; takes about 17 minutes on NWmpp1)*

61. *CSS_test*

V. Customization

62. *Customize the Nodes*

Perform Additional Node Customization (Step 44, Page 65 PSSP Installation Guide)

a) Note: When PSSP installs a node, it uses the AIX sysdumpdev -e command to estimate the size of the dump for the node. PSSP creates a dump logical volume that is approximately 10 percent larger than the estimated dump size, and makes that logical volume the primary dump device. However, you may find that this dump device is not large enough to contain an entire dump due to large processes or applications running on your node. Once your node is up and running, use:
   (1) **sysdumpdev -e** To get the estimated size of the node's dump
   (2) **sysdumpdev -l** To find the name of the primary dump device
   (3) lslv To list the amount of space available in the primary dump device
   (4) **extendlv** To expand the size of the dump logical volume if the estimated dump space is greater than the dump space available

b) On NWMPP1 make sure to "fan-out" (dsh -af 32) as a maximum

c) This will only work if staged over the switch. Otherwise you should use a "fan-out" of 8.

63. *Setup correct routes from CWS to the rest of the node*

a) Default route for NWTEST from nodes to the CWS:

dsh -a "route add default 198.129.163.65" This route will be subsequently added by /etc/rc.local which executes /etc/routes

b) Turn ipforwarding on for node 1:

dsh -w t01n01 "/usr/sbin/no -o ipforwarding=1"

c) Establish a route from the CWS to the switch interface on node 1 (NWTEST example):

host t01s01 198.129.163.130
host t01n01 198.129.163.66
On the CWS type '**smitty mkroute**'
Destination TYPE *net*
Destination *198.129.163.130*
Gateway *198.129.163.66*
Netmask *255.255.255.192*

64. *mount systems on all of the nodes*

a) **dsh -a mkdir /systems**
b) **dsh -a mount prophet:/systems /systems**

65. *Loadleveler Configuration for the control workstation*

Decide on directories for LoadLeveler. Use the information below for the location of the home and local directories. (3.1 Step 1: Decide on Directories for LoadLeveler - LoadLeveler Installation MEMO)

a) Create /loadl filesystem at ~100MB's (NWTEST) on the cw; 434176 MB's (NWMPP1) (*"customize" has already created this filesystem on the nodes check the size of /loadl on the nodes*)
b) **mount /loadl**
c) create loadl userid and group with uid 204 and gid 11 , home = /loadl

(1) **mkgroup -'a' id='11' loadl**
(2) **mkuser -a id='204' pgrp='loadl' home='/loadl' loadl**
(3) Add loadl as a kerberos principal -
    (a) **kadmin**
    (b) **ank loadl** (ank = add_new_key)
(4) **telnet cw**
(5) login as loadl and change passwd
(6) **/usr/lpp/ssp/kerberos/bin/k4init loadl**
(7) exit back to root
(8) run dinner to update the password file on the nodes
(9) **chown loadl.loadl /loadl**
(10) Run the Installation Script ***llinit.***
    http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLIV2R1/3.6
    Step 6: Run the Installation Script llinit - LoadLeveler Installation MEMO) llinit does the following:

- Creates the LoadL_config file and the LoadL_admin file, if they don't already exist.
- Copies the LoadL_admin and the LoadL_config files from the release directory (in the samples subdirectory) into the home directory of loadl. Note: These files are a common resource for all the machines in the LoadLeveler cluster, and therefore must be made accessible to all members of the LoadLeveler pool.
- Sets local directory with permissions set to 775, 700, and 1777, respectively.
- Changing the Location of Directories: You can change the locations of these directories by changing the associated paths in the global configuration file. The global configuration file must reside in loadl's home directory or the location specified in /etc/LoadL.cfg. For example, if you want to move the log, spool, and execute directory from /loadl into /tmp/loadl, with appropriate permissions set, you can do so but you must create /tmp/loadl/spool, /tmp/loadl/execute, and /tmp/loadl/log in /tmp/loadl or LoadLeveler will not start up.
- Copies the LoadL_config.local file from the release directory (in the samples subdirectory) into the local directory.
- Creates symbolic links from the loadl home directory to the spool, execute, and log subdirectories and the LoadL_config.local file in the local directory (if home and local directories are not identical).

(11) You must perform this step using the LoadLeveler user ID. These instructions use loadl as this ID. To switch to the loadl ID, enter the following***:***
(12) su - loadl
(13) Ensure that your HOME environment variable is set to loadl's home directory**. -** ***echo $HOME***
(14) cd /usr/lpp/LoadL/full/bin
(15) Check to make sure bin does not exist in /loadl. Enter the llinit command. For example, to run the llinit command with a local directory of /loadl, a release directory of /usr/lpp/LoadL/full, and a central manager named cw3.nwtest.emsl.pnl.gov, enter the following:
./llinit -local /loadl -release /usr/lpp/LoadL/full -cm cw3.nwtest.emsl.pnl.gov
(16) Make sure to Update the PATH Environment Variable.

66. *Run df -k on node and change filesystem sizes to something reasonable*

a) NWTEST, NWMPP1

/dev/hd4	20480	10172	51%	1793	15% /
/dev/hd2	847872	14708	99%	26936	13% /usr
/dev/hd9var	16384	11724	29%	346	9% /var
/dev/hd3	57344	55440	4%	34	1% /tmp
/dev/hd1	4096	3868	6%	38	4% /home

67. *Setup WCOLL's*

a) Run /systems/cfg/wcoll/copy_<machine_name>_wcolls (i.e. /systems/cfg/wcoll/copy_nwtest_wcolls)
b) Check to make sure the WCOLL's are current

68. *Run customize*

a) What does "*customize/customize.local*" do?
(1) The *customize* script is intended to customize a single node or a set of nodes (e.g. an entire frame or machine dictated by setting WCOLL) after the vanilla install. It is to be run from the control workstation.
(2) The "*customize.local*" script is copied to each node by the *customize script* and is dependent on the "*gender*" file which exists in /systems/cfg/etc/gender and is copied to /etc/gender on each node. The *customize, customize.local* are in */systems/bin/*. The *gender* file must be correct in order for *customize.local* to have a chance at succeeding as it provides the "personality" for each node. The *gender* file distinguishes between node types i.e., "login", "compute", "system", "gpfs", etc. The *customize.local* script attempts to setup your node environment based on how you choose to customize it. This customization includes but is not limited to:
(a) Installing additional license program products (lpps)
(b) FILESYSTEMS
(c) SYSTEM TUNING
(d) NETWORK CUSTOMIZATION
(e) SECURITY CUSTOMIZATION
(f) STATISTICS AND MONITORING
(g) LVM - PHYSICAL VOLUMES/STRIPING/VOLUME GROUP
(h) SPAGING SPACE
(i) SCRATCH
(j) etc.
b) Edit */systems/bin/customize.local* as necessary
c) Make sure "loadl" user and group are created
d) Test customize on a single node by creating a WCOLL called *work*. Then run **/systems/bin/customize work**. If this runs, then do (e)
e) /systems/bin/customize everyone

69. *Configure* "Node Locked Licenses". *Check the compiler documentation for licensing requirements For "node locked" licenses, do the following:*

a) Make sure the following *Program Products* are installed: /systems/lpp/vac44, /systems/lpp/vacpp40, /systems/lpp/aix432/X11.Dt
b) /var/ifor/i4cfg -list *Do this to get license information if working* on CWS and on nodes dsh -a "/var/ifor/i4cfg -list"

c) On the CWS you can run **/systems/bin/resp_lic | /var/ifor/i4cfg -script** and then **/systems/bin/config.lic.**
   - Or **/var/ifor/i4cfg -script** and then answer the questions as follows
     (1) <2>Nodelock License Server
     (2) NO Disable Remote
     (3) <2> Direct Binding
     (4) NO IP Port Number
     (5) <1> Default
     (6) Enter
     (7) NO Modify List
     (8) YES Autostart
     (9) YES Continue
     (10) YES Start Now

d) Then run /systems/bin/config.lic (dsh the command to the nodes)

e) Then run /usr/vac/bin/replaceCSET (dsh the command to the nodes) # this sets up the links. To test: touch foo.c; cc foo.c (dsh the commands to the nodes)

- To manually configure after /var/ifor/i4cfg -script, run /var/ifor/i4cfg -list # should get license information and daemon
- Edit /usr/vac/c44aix_cn.lic as follows:
- replace *"number_of_lic"* with desired concurrent compiles (less than or equal to the legal number of site licenses) *NWTEST = 32*
- replace "admin_name" with "root".
- insert "/var/ifor/" in front of the first line.
- Run the first line from the command line, should look like the following:
  /var/ifor/i4blt -a -v "'IBM Software Solutions Toronto' 5da54a553b4c.02.09.15.31.05.00.00.00 p9gb3ycs6ydpw" -p "'C for AIX' '4.4.cn' 9n688vx858xgap99xv4b9ry7j6227d22 " -T 32 -R root"
- **/usr/vac/bin/replaceCSET** # this sets up the links
- touch foo.c
- cc foo.c # the file will be empty, the compiler will work.

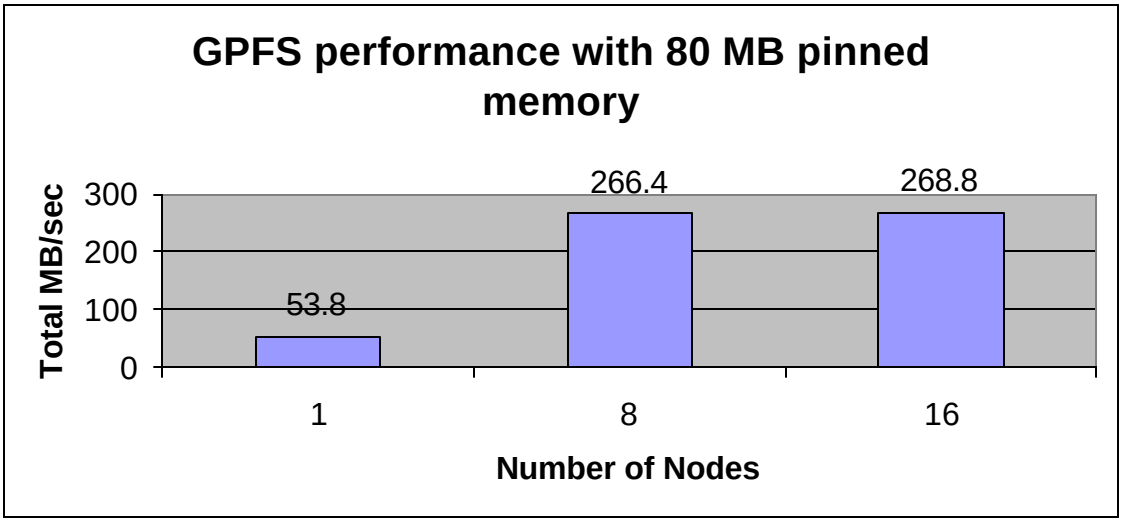
69. *Configuration Management Considerations (*PSSP 3.1 README FIRST)

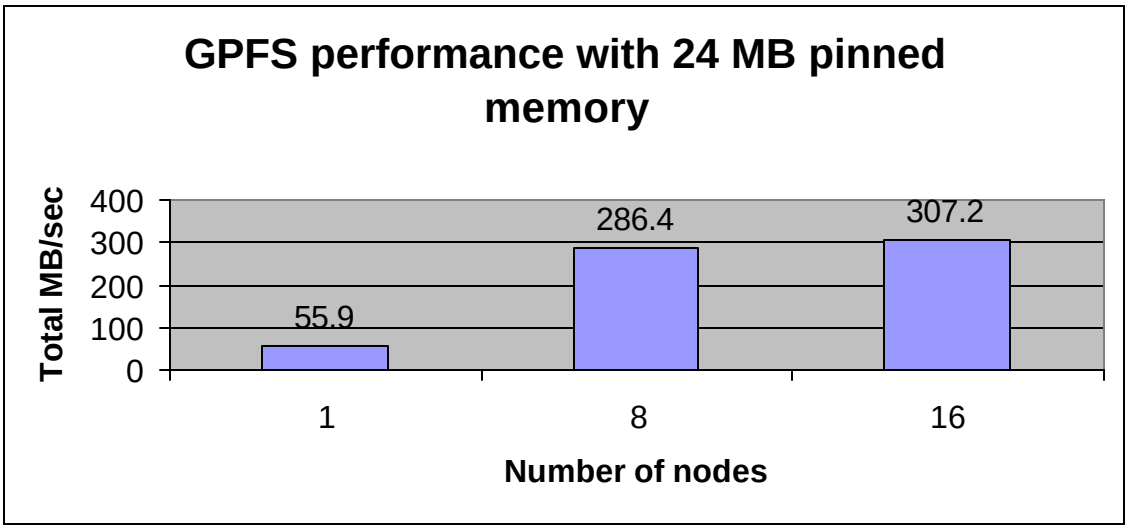
a) To stop configuration manager errors from occurring in the future, if you do not have a graphics adapter in your SP, you should run the following:

(1) rmdev -dl gxme0
(2) rmdev -dl rcm0
(3) rmdev -dl lft0 (if it exists)
(4) /usr/lib/instl/sm_inst installp_cmd -u -f'devices.graphics.com' '-g'

70. *Configure GPFS*

a) Introduction
The following pages will deal exclusively with installing GPFS 1.2 at PNNL on 3 different IBM SP's. These procedures have been tested at our site and are not guaranteed to work elsewhere. We ran performance tests on our 512 node Micro Channel Adapter machine and determined that a large amout of pinned memory allocated to GPFS did not improve our performance. This was due to the MCA SSA Adapters and the type of 2.2GB SSA Disks (See Graphs below).





Assumptions:

80MB = 64MB page pool + 16MB mallocsize
24MB = 20MB pagepool + 4MB mallocsize
All results reflect single runs.
Some tests were performed twice to verify consistency.
All tests run against GPFS1 during normal production.
Read results were ~45%-70% faster/node when recorded.

b) Installation Basics - 512 Node SP

1) Preparation

- Installed software on cw1 "**lslpp -l mmfs***"

```
Fileset                      Level  State      Description
----------------------------------------------------------------------------
Path: /usr/lib/objrepos
mmfs.base.cmds             3.1.0.6  COMMITTED  GPFS File Manager Commands
mmfs.base.rte              3.1.0.7  COMMITTED  GPFS File Manager
mmfs.gpfs.rte              1.2.0.5  COMMITTED  GPFS File Manager
mmfs.msg.en_US             3.1.0.4  COMMITTED  GPFS Server Messages - U.S.
                                                English
mmfs.util.smit             3.1.0.1  COMMITTED  GPFS Server SMIT Panels
```

```
Path: /etc/objrepos
mmfs.base.rte             3.1.0.7  COMMITTED  GPFS File Manager
mmfs.gpfs.rte             1.2.0.0  COMMITTED  GPFS File Manager

Path: /usr/share/lib/objrepos
mmfs.gpfsdocs.data        3.1.0.0  COMMITTED  GPFS Server Man Pages and
                                              Documentation
mmfs.man.en_US.data       3.1.0.0  COMMITTED  GPFS Server Man Pages - U.S.
                                              English
```

- Installed software on every node **"dsh -av lslpp -l mmfs* | dshbak -c > /tmp/lpps.gpfs"**

```
Fileset                      Level  State      Description
----------------------------------------------------------------------------
Path: /usr/lib/objrepos
mmfs.base.cmds            3.1.0.6  COMMITTED  GPFS File Manager Commands
mmfs.base.rte             3.1.0.7  COMMITTED  GPFS File Manager
mmfs.gpfs.rte             1.2.0.5  COMMITTED  GPFS File Manager
mmfs.msg.en_US            3.1.0.4  COMMITTED  GPFS Server Messages - U.S.
                                              English
mmfs.util.smit            3.1.0.1  COMMITTED  GPFS Server SMIT Panels

Path: /etc/objrepos
mmfs.base.rte             3.1.0.7  COMMITTED  GPFS File Manager
mmfs.gpfs.rte             1.2.0.0  COMMITTED  GPFS File Manager
```

- SP Switch Tuning
  "dsh -av lsattr -El css0 | dshbak -c > /tmp/switch.out"
  In our case we want the spoolsize and rpoolsize set to 16MB each.

  If any entry is wrong, run **"/usr/lpp/ssp/css/chgcss -l css0 -a spoolsize=16777216 -a rpoolsize=16777216"** on that node and reboot the node to make it take affect.

  We also enter this command in /etc/rc.net at the end Part I so a boot will make it take affect.

```
-------------------------------------------------------------
bus_mem_addr   0x04000000 Bus memory address        False
int_level      0xb        Bus interrupt level       False
int_priority   3          Interrupt priority        False
dma_lvl        9          DMA arbitration level     False
spoolsize      16777216   Size of IP send buffer    True
rpoolsize      16777216   Size of IP receive buffer True
adapter_status css_ready  Configuration status      False
```

  Run "**SDRGetObjects TS_Config**" to get the following:

```
   Frequency    Sensitivity  Run_FixPri   FixPri_Value Log_Length   Pinning
       2            8            1           38         5000         Text
```

  If the current system does not reflect these options run "**SDRChangeAttrValues TS_Config Sensitivity=6**".
  If you make a change, refresh hats and hags (topology services) by:
  "dsh -av refresh -s hats"
  From cw1, **"refresh -s hats.cw1"**
  "dsh -av refresh -a hags"
  From cw1, **"refresh -s hags.cw1"**
- Configure sysctl
  "dsh -av cat /etc/sysctl.vsd.acl | dshbak -c > /tmp/sysctl.out"

  This entry should look the same on cw1 also.

```
------------------------------------------------------------
#acl#

# These are the users that can issue sysctl_vsdXXX command on
this node
# Name must have a Kerberos name format which defines
user@realm
# Please check your security administrator to fill in correct
realm name
# you may find realm name from /etc/krb.conf

# _PRINCIPAL root@PPD.POK.IBM.COM
# _PRINCIPAL root.admin@PPD.POK.IBM.COM
# _PRINCIPAL rcmd@PPD.POK.IBM.COM
# _PRINCIPAL userid@PPD.POK.IBM.COM
_PRINCIPAL root@NWMPP1.EMSL.PNL.GOV
_PRINCIPAL root.admin@NWMPP1.EMSL.PNL.GOV
_PRINCIPAL rcmd@NWMPP1.EMSL.PNL.GOV
_PRINCIPAL root.SPbgAdm
```

dsh -av "cat /etc/sysctl.mmcmd.acl" | dshbak -c > /tmp/sysctl.mmcmd.out

This entry should also look the same on cw1

```
-----------------------------------------------------
#acl#
# These are the users that can issue multinode mmfs commands
through sysctl:
#_PRINCIPAL root.admin@PPD.POK.IBM.COM
#_PRINCIPAL root@PPD.POK.IBM.COM
_PRINCIPAL root@NWMPP1.EMSL.PNL.GOV
_PRINCIPAL root.admin@NWMPP1.EMSL.PNL.GOV
_PRINCIPAL rcmd@NWMPP1.EMSL.PNL.GOV
```

dsh -av "cat /etc/sysctl.acl" | dshbak -c > /tmp/sysctl.acl.out

This entry should also look the same on cw1

```
------------------------------------------------------------
#acl#
# This sample acl file contains commented out lines for a
principal
# and an acl file.
#_PRINCIPAL root.admin@HPSSL.KGN.IBM.COM
#_ACL_FILE /etc/mycmd.sysctl.acl
_PRINCIPAL root@NWMPP1.EMSL.PNL.GOV
_PRINCIPAL root.admin@NWMPP1.EMSL.PNL.GOV
_PRINCIPAL rcmd@NWMPP1.EMSL.PNL.GOV
------------------------------------------------------------
```

dsh -av "sysctl svcrestart"
"sysctl svcrestart" on cw1 also
Test the sysctl authorization by:
k4list
sysctl whoami
sysctl sysctl_vsdcheck
vsdsklst -a

vsdsklst -n <pick-a-node-number>
dsh -av "sysctl mmremote single", you'll want to see an OK at each node response.

If there's an authorization problem:

tail -4 /etc/sysctl.conf, make sure that the following include files are there:

```
# Include VSD sysctl_vsd.cmd commands
include /usr/lpp/csd/sysctl/sysctl_vsd.cmds
# Include mmfs sysctl commands
include /usr/lpp/mmfs/bin/mmcmdsysctl
dsh -av "stopsrc -s sysctld"
dsh -av "startsrc -s sysctld"
```

2) Steps in creating RVSD's

Now that we have all the software in place and acl's have been configured and tested it's time to create the Virtual Shared Disks and the Recoverable Virtual Shared Disks. The RVSD's are a totally separate activity from GPFS, but GPFS is completely dependent on having properly configured RVSD's. There are several steps to creating the RVSD's:

- "vsdnode" determines which nodes are in the group, both servers and clients.
- Create a "dummy" VSD on any node just to get some subsystems running, this is later removed.
- Determine the "shortest path" between a disk server node and it's SSA disks. We can view this using "diag" and also IBM's tool "maymap". We will use twintailing to improve availability of the SSA disks. Twintailing allows us to completely loose a disk server node but automatically shift the disks over to the designated node for serving.
- Using the list created, run "createvsd" with appropriate parameters for our system.

3) Create the RVSD's (VSD's)

- We enable the VSD's on all 512 nodes (just to make sure), allowing 2 Buddy Buffers on all Client Nodes and 64 Buddy Buffers on all Server Nodes. The Server Nodes operate the SSA Disk Drives and are dedicated to this purpose, so pinning memory does not harm other needs.

```
/usr/lpp/csd/bin/vsdnode \
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 50 51 52 53 54 55 56 57 58 59 60 61
62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89
90 91 92 93 94 95 96 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113
114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134
135 136 137 138 139 140 141 142 143 144 146 147 148 149 150 151 152 153 154 155 156
157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177
178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 194 195 196 197 198 199
200 201 202 203 204 205 206 207 208 215 217 219 221 223 231 233 235 237 239 242 243
244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264
265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285
286 287 288 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307
308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328
329 330 331 332 333 334 335 336 338 339 340 341 342 343 344 345 346 347 348 349 350
351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
372 373 374 375 376 377 378 379 380 381 382 383 384 387 388 389 390 391 392 393 394
395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415
```

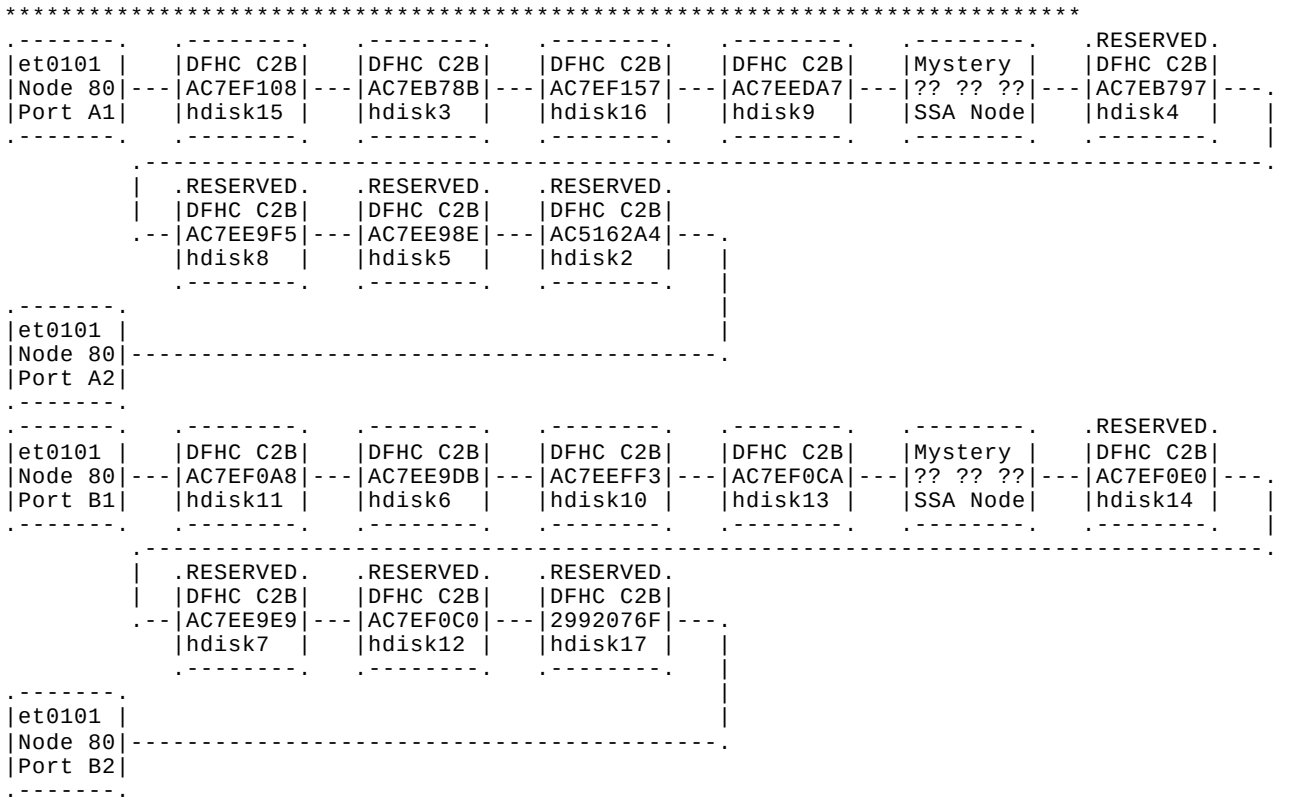
```
416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 434 435 436 437
438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458
459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479
480 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501
502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522
523 524 525 526 527 528 \
css0 64 256 256 48 4096 262144 2 61440

/usr/lpp/csd/bin/vsdnode \
1 49 97 145 193 241 289 337 385 386 433 481 209 211 213 225 227 229 \
css0 64 256 256 48 4096 262144 64 61440
```

To verify nodes were properly set up, run "**vsdatalst -n**", a sample follows:

VSD Node Information

node number	host_name	VSD adapter	IP packet size	Initial cache buffers	Maximum cache buffers	VSD request count	rw request count	Buddy Buffer minimum size	Buddy Buffer maximum size	Buddy Buffer: # maxbufs
1	cef1.nwmpp1.ems	css0	61440	64	256	256	48	4096	262144	64
2	et0102.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
3	et0103.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
4	et0104.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
5	et0105.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
6	et0106.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
7	et0107.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
8	et0108.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
9	et0109.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2
10	et0110.nwmpp1.e	css0	61440	64	256	256	48	4096	262144	2

- Create a "dummy" VSD on node 1 (et0101) to activate some subsystems, we will remove it later.
  From cw1 "**/usr/lpp/csd/bin/createvsd -n 1 -g rootvg -s 8 -c 1 -T 8 -o nocache**"
  From et0101 "**vsdatalst -v**" to view the new "dummy" vsd and confirm it's name is "vsd1n1".
  From cw1 "**dsh -av /usr/lpp/csd/bin/cfgvsd vsd1n1**"
  From et0101 "**/usr/lpp/csd/bin/lsvsd -l**", make sure "ACT" appears in "State" column.
  From et0101 "**vsdvts vsd1n1**" to test the dummy vsd, "Continue? YES"
  To cleanup this dummy vsd:
  From cw1 "**dsh -av suspendvsd vsd1n1**"
  From cw1 "**dsh -av stopvsd vsd1n1**"
  From cw1 "**dsh -av ucfgvsd vsd1n1**"
  From cw1 "**removevsd -v vsd1n1 -f**" to force removal of this dummy vsd.
  From cw1 "**dsh -av lsvsd -l**", you should get errors from every node.
- Create the VSD's for /gpfs1 first.
  This includes 96 2.2GB SSA Disks spread over 12 dedicated thin node servers. Each server node can see 16 SSA Disks but will only actively use the 8 assigned to that node. These 8 need to be the 8 closest to that node's SSA Adapter on the A1 Port. Doing that will prevent "cross talking" across the SSA cable which would slow down other disks. You can choose to use A1, B1, A2 or B2 as long as the other nodes are twintailed. There are 2 ways to determine which 8 disks are to be used, maymap and diag. I placed "maymap" into /systems/gpfstest/maymap/maymap and a sample output appears following:



****

The "Mystery" box is actually the twintail node these disks are shared by, in this case et0401 or node 49. There are options with "maymap" that allow output other than "boxes", running "maymap -lph" gives the following:

****

```
Resrc ID Serial No UID      : A1 A2 B1 B2 HWI Status   Model Code Lvl          Reserved pdisk   hdisk
0301001C 68D79052  AC7EF108 : 00 08 FF FF (0=ST_Good)        DFHC C2B RAMST077          pdisk13 hdisk15
0301001D 68E13432  AC7EB78B : 01 07 FF FF (0=ST_Good)        DFHC C2B RAMST077          pdisk1  hdisk3
03010012 68E02435  AC7EF157 : 02 06 FF FF (0=ST_Good)       DFHC C2B RAMST077          pdisk14 hdisk16
03010013 68D92244  AC7EEDA7 : 03 05 FF FF (0=ST_Good)        DFHC C2B RAMST077          pdisk7  hdisk9
0301000A 68D99481  AC7EB797 : 05 03 FF FF (1=ST_Failed)      DFHC C2B RAMST077 Reserved pdisk2  hdisk4
0301000C 68E06036  AC7EE9F5 : 06 02 FF FF (1=ST_Failed)       DFHC C2B RAMST077 Reserved pdisk6  hdisk8
0301001A 68E09226  AC7EE98E : 07 01 FF FF (1=ST_Failed)      DFHC C2B RAMST077 Reserved pdisk3  hdisk5
03010011 68406694  AC5162A4 : 08 00 FF FF (1=ST_Failed)      DFHC C2B RAMST077 Reserved pdisk0  hdisk2
0301001E 68E09217  AC7EF0A8 : FF FF 00 08 (0=ST_Good)        DFHC C2B RAMST077          pdisk9  hdisk11
0301000E 68D92377  AC7EE9DB : FF FF 01 07 (0=ST_Good)        DFHC C2B RAMST077          pdisk4  hdisk6
0301001B 68D72795  AC7EEFF3 : FF FF 02 06 (0=ST_Good)        DFHC C2B RAMST077          pdisk8  hdisk10
0301000D 68E08786  AC7EF0CA : FF FF 03 05 (0=ST_Good)       DFHC C2B RAMST077          pdisk11 hdisk13
0301000B 68D92270  AC7EF0E0 : FF FF 05 03 (1=ST_Failed)      DFHC C2B RAMST077 Reserved pdisk12 hdisk14
0301000F 68E06617  AC7EE9E9 : FF FF 06 02 (1=ST_Failed)      DFHC C2B RAMST077 Reserved pdisk5  hdisk7
03010010 68D88875  AC7EF0C0 : FF FF 07 01 (1=ST_Failed)      DFHC C2B RAMST077 Reserved pdisk10 hdisk12
0301001F 13A25083  2992076F : FF FF 08 00 (1=ST_Failed)       DFHC C2B RAMST077 Reserved pdisk15 hdisk17
```

****

The column with the "reserved" refers to SSA Disks assigned to this serving node's twintail partner node. If you don't have maymap, enter "diag", "Task Selection (Diagnostics, Advanced Diagnostics, Service Aids, etc.)", "SSA Service Aids", "Link

Verification", "et0101:ssa0 00-03 SSA Adapter" which will display the following:

```
  Physical                          Serial#     Adapter Port
                                                A1  A2  B1  B2      Status

[TOP]
  et0101:pdisk13                    AC7EF108     0   8              Good
  et0101:pdisk1                     AC7EB78B     1   7              Good
  et0101:pdisk14                    AC7EF157     2   6              Good
  et0101:pdisk7                     AC7EEDA7     3   5              Good
  *****                                          4   4
  et0101:pdisk2                     AC7EB797     5   3              Reserved
  et0101:pdisk6                     AC7EE9F5     6   2              Reserved
  et0101:pdisk3                     AC7EE98E     7   1              Reserved
  et0101:pdisk0                     AC5162A4     8   0              Reserved
  et0101:pdisk9                     AC7EF0A8              0   8      Good
  et0101:pdisk4                     AC7EE9DB              1   7      Good
  et0101:pdisk8                     AC7EEFF3              2   6      Good
  et0101:pdisk11                    AC7EF0CA              3   5      Good
  *****                                                   4   4
  et0101:pdisk12                    AC7EF0E0              5   3      Reserved
  et0101:pdisk5                     AC7EE9E9              6   2      Reserved
  et0101:pdisk10                    AC7EF0C0              7   1      Reserved
  et0101:pdisk15                    2992076F              8   0      Reserved
```

This procedure requires that you determine the relationship between "pdisk" and "hdisk" since hdisk is required to create the vsd's and I use the pdisk the give a volume name to the disk which helps in problem determination. Step back a couple screen on the diag till you see "Configuration Verification" and follow that path, it will display the pdisk to hdisk relationship. After all this is determined, I created a file that is executable.

The following is the contents of the createvsd.gpfs_servers.sh file:

```
date #1/49
createvsd -n 1/49:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd -n 1/49:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
createvsd -n 1/49:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
createvsd -n 1/49:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
#
createvsd -n 1/49:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 1/49:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 1/49:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
createvsd -n 1/49:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
date
date #49/1
createvsd -n 49/1:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 49/1:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
createvsd -n 49/1:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
createvsd -n 49/1:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
#
createvsd -n 49/1:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 49/1:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
createvsd -n 49/1:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
createvsd -n 49/1:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
date
date #97/145
createvsd -n 97/145:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd -n 97/145:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 97/145:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
createvsd -n 97/145:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
```

```
#
createvsd -n 97/145:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
createvsd -n 97/145:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 97/145:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
createvsd -n 97/145:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
date #145/97
createvsd -n 145/97:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 145/97:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
createvsd -n 145/97:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 145/97:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
#
createvsd -n 145/97:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
createvsd -n 145/97:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
createvsd -n 145/97:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
createvsd -n 145/97:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
date
date #193/241
createvsd -n 193/241:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 193/241:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 193/241:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
createvsd -n 193/241:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
#
createvsd -n 193/241:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 193/241:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
createvsd -n 193/241:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
createvsd -n 193/241:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
date
date #241/193
createvsd –n 241/193:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
createvsd -n 241/193:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
createvsd -n 241/193:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 241/193:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 241/193:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
createvsd -n 241/193:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 241/193:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
createvsd -n 241/193:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
date
date #289/337
createvsd -n 289/337:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 289/337:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
createvsd -n 289/337:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 289/337:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
#
createvsd -n 289/337:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
createvsd -n 289/337:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 289/337:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 289/337:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
date #337/289
createvsd -n 337/289:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
createvsd -n 337/289:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
createvsd -n 337/289:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
createvsd -n 337/289:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
#
createvsd -n 337/289:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 337/289:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
createvsd -n 337/289:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 337/289:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
date
date #385/386
createvsd -n 385/386:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 385/386:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
createvsd -n 385/386:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 385/386:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
#
createvsd -n 385/386:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
createvsd -n 385/386:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 385/386:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
createvsd -n 385/386:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
date #386/385
createvsd -n 386/385:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 386/385:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
createvsd -n 386/385:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 386/385:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
```

```
#
createvsd -n 386/385:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
createvsd -n 386/385:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 386/385:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd -n 386/385:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
date
date #433/481
createvsd -n 433/481:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 433/481:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
createvsd -n 433/481:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
createvsd -n 433/481:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
#
createvsd -n 433/481:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 433/481:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
createvsd -n 433/481:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 433/481:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
#
date #481/433
createvsd -n 481/433:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 481/433:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
createvsd -n 481/433:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
createvsd -n 481/433:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
#
createvsd -n 481/433:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 481/433:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd -n 481/433:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
createvsd -n 481/433:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
date
```

Every line entry will create output similar to the following:

```
********************************************************************************
+ createvsd -n 1/49:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd: calls Getopts.
createvsd: parsing node_list.
createvsd: call vsd_rollback.
It took about 1 seconds in vsd_rollback.
createvsd: creates task tables.
createvsd: calls sysctl_mkvglv.
It took about 10 seconds in mkvglv.
createvsd: calls sysctl_varyoffvg 15vg on the primary node
cef1.nwmpp1.emsl.pnl.gov
createvsd: calls sysctl_importvg 15vg on the backup node cef4.nwmpp1.emsl.pnl.gov
with 00006186d171b9ff
It took about 28 seconds in importvg.
createvsd: calls sysctl_varyonvg 15vg on the primary node
cef1.nwmpp1.emsl.pnl.gov
createvsd: calls vsdvg.
OK:0:vsdvg -g 15vgn1b49 15vg 1 49
It took about 3 seconds in vsdvg.
createvsd: calls defvsd.
OK:0:defvsd 15lv4n1 15vgn1b49 15v4n1 nocache
It took about 3 seconds in defvsd.
********************************************************************************
```

- Now create the vsd's for /gpfs2.
  This group contains 48 2.2GB SSA Disks twintailed just like the servers for /gpfs1 previously mentioned. Teh following is the contents of createvsd.widenodes.041300:

```
date #209/225
createvsd -n 209/225:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
createvsd -n 209/225:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
createvsd -n 209/225:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 209/225:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
#
createvsd -n 209/225:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 209/225:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 209/225:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
createvsd -n 209/225:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
date
date #225/209
createvsd -n 225/209:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
```

```
createvsd -n 225/209:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 225/209:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 225/209:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
#
createvsd -n 225/209:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
createvsd -n 225/209:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd -n 225/209:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 225/209:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
date
date #211/227
createvsd -n 211/227:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 211/227:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
createvsd -n 211/227:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
createvsd -n 211/227:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
#
createvsd -n 211/227:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 211/227:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 211/227:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
createvsd -n 211/227:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
date
date #227/211
createvsd -n 227/211:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
createvsd -n 227/211:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
createvsd -n 227/211:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
createvsd -n 227/211:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
#
createvsd -n 227/211:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 227/211:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
createvsd -n 227/211:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 227/211:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
date
date #213/229
createvsd -n 213/229:hdisk16/ -s 2148 -g 16vg -v 16v -l 16lv -T 4 -o nocache
createvsd -n 213/229:hdisk14/ -s 2148 -g 14vg -v 14v -l 14lv -T 4 -o nocache
createvsd -n 213/229:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd -n 213/229:hdisk13/ -s 2148 -g 13vg -v 13v -l 13lv -T 4 -o nocache
#
createvsd -n 213/229:hdisk17/ -s 2148 -g 17vg -v 17v -l 17lv -T 4 -o nocache
createvsd -n 213/229:hdisk2/  -s 2148 -g  2vg -v 2v  -l 2lv  -T 4 -o nocache
createvsd -n 213/229:hdisk4/  -s 2148 -g  4vg -v 4v  -l 4lv  -T 4 -o nocache
createvsd -n 213/229:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4 -o nocache
date
date #229/213
createvsd -n 229/213:hdisk10/ -s 2148 -g 10vg -v 10v -l 10lv -T 4 -o nocache
createvsd -n 229/213:hdisk9/  -s 2148 -g  9vg -v  9v -l 9lv  -T 4 -o nocache
createvsd -n 229/213:hdisk12/ -s 2148 -g 12vg -v 12v -l 12lv -T 4 -o nocache
createvsd -n 229/213:hdisk11/ -s 2148 -g 11vg -v 11v -l 11lv -T 4 -o nocache
#
createvsd -n 229/213:hdisk7/  -s 2148 -g  7vg -v  7v -l  7lv -T 4 -o nocache
createvsd -n 229/213:hdisk5/  -s 2148 -g  5vg -v 5v  -l 5lv  -T 4 -o nocache
createvsd -n 229/213:hdisk6/  -s 2148 -g  6vg -v  6v -l  6lv -T 4 -o nocache
createvsd -n 229/213:hdisk8/  -s 2148 -g  8vg -v 8v  -l 8lv  -T 4 -o nocache
date
```

Again create a "for" loop and run every line, redirect the output to a file for review to assure that every line was completed sucessfully.

Configure the newly created VSD's by running "**cfgvsd -a**" on cw1, then "**lsvsd -l**" from et0101 to verify they are all there and configured. Start these VSD's using **"ha.vsd start"**, then confirm they are active by doing **"lsvsd –l"**.

4) Creating GPFS File Systems

- ### Stripe Group Manager

Knowing where the "Stripe Group Manager" and the "Cluster Config Manager" are located is very valuable, we created a

file listing all the VSD Server Nodes in it so that when the managers are assigned they will be picked only from the dedicated servers. This prevents one of these important services from runnig on just any node which might get overwhelmed by a user's job. This preference file is located in /var/mmfs/etc/cluster.preferences on every node and for NWMPP1 looks like the following:

```
sw0101.nwmpp1.emsl.pnl.gov
sw0401.nwmpp1.emsl.pnl.gov
sw0701.nwmpp1.emsl.pnl.gov
sw1001.nwmpp1.emsl.pnl.gov
sw1301.nwmpp1.emsl.pnl.gov
sw1601.nwmpp1.emsl.pnl.gov
sw1901.nwmpp1.emsl.pnl.gov
sw2201.nwmpp1.emsl.pnl.gov
sw2501.nwmpp1.emsl.pnl.gov
sw2801.nwmpp1.emsl.pnl.gov
sw3101.nwmpp1.emsl.pnl.gov
sw1401.nwmpp1.emsl.pnl.gov
sw1403.nwmpp1.emsl.pnl.gov
sw1405.nwmpp1.emsl.pnl.gov
sw1501.nwmpp1.emsl.pnl.gov
sw1503.nwmpp1.emsl.pnl.gov
sw1505.nwmpp1.emsl.pnl.gov
```

To find which nodes are the managers, from any node run "**mmfsadm dump cfgmgr**" and look for: "Cluster configuration manager is 198.129.183.34 (other node)"

"Stripe groups managed by this node:" Do a host on the IP address to get "sw2801.nwmpp1.emsl.pnl.gov is 198.129.183.34, Aliases: sw2801, m28s01" in this case, login there and run "**mmfsadm dump cfgmgr**" again. The results on et2801 are:

Cluster configuration manager is 198.129.183.34 (this node)

Assigned stripe group managers: "gpfs1" SG mgr 198.129.183.82

Appointed at 957477640.959714304; done recovery=true "gpfs2" SG mgr 198.129.181.98

Appointed at 957477640.958670848; done recovery=true Stripe groups managed by this node: (none)

There is only 1 Cluster Manager per system but a Stripe Group Manager for every gpfs file system. When there is a problem with gpfs, a solution that worked for us was to stop and start hats/hags on the Cluster Manager node, that caused a reset and the Manager moved to another node in the preference list. Cleared up a lot of strange symptoms.

- Configure and start the GPFS File System
  login to et0101
  "**k4init root.admin**"
  "**dsh -av -a startsrc -s mmfs**"
  "**/usr/lpp/mmfs/bin/mmconfig -a -m 4M -M 200 -p 20M**" will configure all nodes with mallocsize of 4M, MaxFilesToCache=200 and pagepool=20M, while the subsystem will not automatically

start-up, we like to manually start mmfs due to resource control. To verify, **cat /var/mmfs/etc/mmfs.cfg** and view the options.

- Create File Descriptions that MMFS needs
  Call it /var/mmfs/scratchdesc for /gpfs1 and /var/mmfs/scratchdesc2 for /gpfs2.

  Place these files in et0101 which is where the mmfs commands will run from. This file consists of all the VSD names found in the first column of the command "**vsdatalst -v**" that are to serve a particular gpfs filesystem. We use 2 filesystems, /gpfs1 with 96 2.2 GB Disks and /gpfs2 with 48 2.2 GB Disks, so only include the VSD names you want to serve that particular file system. An example of the first couple of VSD names for /gpfs1 follows (just use the first column):

VSD name	logical volume	Global Volume Group	minor#	option	size_in_MB
10v4n1	10lv4n1	10vgn1b49	55	nocache	2148
11v4n1	11lv4n1	11vgn1b49	53	nocache	2148
13v4n1	13lv4n1	13vgn1b49	56	nocache	2148
15v4n1	15lv4n1	15vgn1b49	49	nocache	2148
16v4n1	16lv4n1	16vgn1b49	51	nocache	2148
3v15n1	3lv15n1	3vgn1b49	50	nocache	2148
6v15n1	6lv15n1	6vgn1b49	54	nocache	2148
9v4n1	9lv4n1	9vgn1b49	52	nocache	2148

- To actually create the /gpfs1 filesystem, run "**/usr/lpp/mmfs/bin/mmcrfs /gpfs1 /dev/gpfs1 -F /var/mmfs/scratchdesc -A no -B 256K -i 4K -I 16K -n 512 -N 128K -v no -R 1 -M 1 -r 1 -m 1 -Q no -s roundRobin**"

- After completion, run "**mmlsfs /dev/gpfs1**" from et0101 to get:

```
flag value          description
---- -------------- ------------------------------------------------------
 -s  roundRobin     Stripe method
 -f  8192           Minimum fragment size in bytes
 -i  4096           Inode size in bytes
 -I  16384          Indirect block size in bytes
 -m  1              Default number of metadata replicas
 -M  1              Maximum number of metadata replicas
 -r  1              Default number of data replicas
 -R  1              Maximum number of data replicas
 -a  1649664        Estimated average file size
 -n  512            Estimated number of nodes that will mount file system
 -B  262144         Block size
 -Q  none           Quotas enforced
 -F  135168         Maximum number of inodes
 -V  2              File system version. Highest supported version: 2
 -d
15v4n1;3v15n1;16v4n1;9v4n1;11v4n1;6v15n1;10v4n1;13v4n1;4v15n49;8v4n49;5v15n49;2v14n49
;14v4n49;7v15n49;12v4n49;17v4n49;15v5n97;4v16n97;5v16n97;14v5n97;7v16n97;11v5n97;10v5
n97;13v5n97;17v5n145;3v16n145;2v15n145;6v16n145;8v5n145;12v5n145;16v5n145;9v5n145;6v2
0n193;2v20n193;3v20n193;12v9n193;17v9n193;10v9n193;5v20n193;7v20n193;8v9n241;9v9n241;
11v9n241;4v20n241;16v9n241;14v9n241;13v9n241;15v9n241;4v17n289;9v6n289;6v17n289;7v17n
289;12v6n289;14v6n289;17v6n289;3v17n289;10v6n337;13v6n337;8v6n337;16v6n337;11v6n337;5
v17n337;2v17n337;15v6n337;4v18n385;3v18n385;2v18n385;7v18n385;8v7n385;17v7n385;13v7n3
85;5v18n385;6v18n386;9v7n386;14v7n386;10v7n386;12v7n386;11v7n386;15v7n386;16v7n386;6v
19n433;13v8n433;16v8n433;7v19n433;17v8n433;10v8n433;14v8n433;3v19n433;2v19n481;8v8n48
1;12v8n481;4v19n481;11v8n481;15v8n481;9v8n481;5v19n481  Disks in file system
 -C  1              Configuration identifier
```

- To actually create the /gpfs2 filesystem, run: "**/usr/lpp/mmfs/bin/mmcrfs /gpfs2 /dev/gpfs2 -F /var/mmfs/scratchdesc2 -A no -B 256K -i 4K -I 16K -n 512 -N 128K -v no -R 1 -M 1 -r 1 -m 1 -Q no -s roundRobin**"
- After completion, run "**mmlsfs /dev/gpfs2**" from et0101 to get:

```
flag value          description
---- -------------- ------------------------------------------------------
 -s  roundRobin     Stripe method
 -f  8192           Minimum fragment size in bytes
 -i  4096           Inode size in bytes
 -I  16384          Indirect block size in bytes
 -m  1              Default number of metadata replicas
 -M  1              Maximum number of metadata replicas
 -r  1              Default number of data replicas
 -R  1              Maximum number of data replicas
 -a  824832         Estimated average file size
 -n  512            Estimated number of nodes that will mount file system
 -B  262144         Block size
 -Q  none           Quotas enforced
 -F  132096         Maximum number of inodes
 -V  2              File system version. Highest supported version: 2
 -d
9v10n209;13v10n209;6v21n209;10v10n209;17v10n209;11v10n209;7v111n209;5v21n209;3v111n22
5;2v21n225;4v21n225;12v10n225;16v10n225;15v10n225;14v10n225;8v10n225;14v11n211;8v11n2
11;7v112n211;15v11n211;6v111n211;2v111n211;13v11n211;5v112n211;11v11n227;3v112n227;12
v11n227;16v11n227;17v11n227;10v11n227;4v112n227;9v11n227;16v12n213;14v12n213;15v12n21
3;13v12n213;17v12n213;2v112n213;4v113n213;3v113n213;10v12n229;9v12n229;12v12n229;11v1
2n229;7v113n229;5v113n229;6v113n229;8v12n229  Disks in file system -C  1
Configuration identifier
```

- Mount the new filesystems "**dsh -av mount /gpfs1**", "**dsh -av mount /gpfs2**"
- We use gpfs for scratch space for our users, ~200GB for /gpfs1 and ~100GB for /gpfs2. To do this requires that user directories be created on the files space, one directory for every user.

Run "**/usr/local/sbin/add_all_gpfs1_users**" from et0101 then run "**/usr/local/sbin/add_all_gpfs2_users**" to create the appropriatedirectories. These are "locally" created Korn Shell scripts that grab users from the "password" file and then add directories owned by the users to GPFS.

c) Installation Basics 8 Node SP NWTEST

1) Preparation

- Installed Software on cw3 "**lslpp –l mmfs***"

```
Fileset                      Level  State      Description
  ----------------------------------------------------------------------------
Path: /usr/lib/objrepos
  mmfs.base.cmds             3.1.0.6  COMMITTED  GPFS File Manager Commands
  mmfs.base.rte              3.1.0.7  COMMITTED  GPFS File Manager
  mmfs.gpfs.rte              1.2.0.5  COMMITTED  GPFS File Manager
  mmfs.msg.en_US             3.1.0.4  COMMITTED  GPFS Server Messages - U.S.
                                                 English
  mmfs.util.smit             3.1.0.1  COMMITTED  GPFS Server SMIT Panels

Path: /etc/objrepos
  mmfs.base.rte              3.1.0.7  COMMITTED  GPFS File Manager
  mmfs.gpfs.rte              1.2.0.0  COMMITTED  GPFS File Manager

Path: /usr/share/lib/objrepos
  mmfs.gpfsdocs.data         3.1.0.1  COMMITTED  GPFS Server Man Pages and
                                                 Documentation
  mmfs.man.en_US.data        3.1.0.0  COMMITTED  GPFS Server Man Pages - U.S.
                                                 English
```

- Installed software on every node "**dsh -av lslpp -l mmfs* | dshbak -c > /tmp/lpps.gpfs**"

```
Fileset                      Level  State      Description
  ----------------------------------------------------------------------------
Path: /usr/lib/objrepos
  mmfs.base.cmds             3.1.0.6  COMMITTED  GPFS File Manager Commands
  mmfs.base.rte              3.1.0.7  COMMITTED  GPFS File Manager
  mmfs.gpfs.rte              1.2.0.5  COMMITTED  GPFS File Manager
  mmfs.util.smit             3.1.0.1  COMMITTED  GPFS Server SMIT Panels

Path: /etc/objrepos
  mmfs.base.rte              3.1.0.7  COMMITTED  GPFS File Manager
  mmfs.gpfs.rte              1.2.0.0   COMMITTED  GPFS File Manage
```

- SP Switch Tuning "**dsh -av lsattr -El css0 | dshbak -c > /tmp/switch.out**"

  In our case we want the spoolsize and rpoolsize set to 16MB each. If any entry is wrong, run "**/usr/lpp/ssp/css/chgcss -l css0 -a spoolsize=16777216 -a rpoolsize=16777216**" on that node and reboot the node to make it take affect. We also enter this command in /etc/rc.net at the end Part I so a boot will make it take affect.

```
  ----------------------------------------------------------
  bus_mem_addr   0x04000000 Bus memory address        False
  int_level      0xb        Bus interrupt level       False
  int_priority   3          Interrupt priority        False
  dma_lvl        9          DMA arbitration level     False
  spoolsize      16777216   Size of IP send buffer    True
  rpoolsize      16777216   Size of IP receive buffer True
  adapter_status css_ready  Configuration status      False
```

  Run "SDRGetObjects TS_Config" to get the following:

```
  Frequency    Sensitivity  Run_FixPri   FixPri_Value Log_Length   Pinning
      1            6            1           38         5000 ""
```

  These are the defaults and are sufficient for a small 8 node system.

- Configure sysctl "**dsh -av cat /etc/sysctl.vsd.acl | dshbak -c > /tmp/sysctl.out**". This entry should look the same on cw3 also.

```
--------------------------------------------------------------------------------
#acl#

# These are the users that can issue sysctl_vsdXXX command on this node
# Name must have a Kerberos name format which defines user@realm
# Please check your security administrator to fill in correct realm name
# you may find realm name from /etc/krb.conf

# _PRINCIPAL root@PPD.POK.IBM.COM
# _PRINCIPAL root.admin@PPD.POK.IBM.COM
# _PRINCIPAL rcmd@PPD.POK.IBM.COM
# _PRINCIPAL userid@PPD.POK.IBM.COM
_PRINCIPAL root@NWTEST.EMSL.PNL.GOV
_PRINCIPAL root.admin@NWTEST.EMSL.PNL.GOV
_PRINCIPAL rcmd@NWTEST.EMSL.PNL.GOV
_PRINCIPAL root.SPbgAdm
--------------------------------------------------------------------------------
```

dsh -av "cat /etc/sysctl.mmcmd.acl" | dshbak -c > /tmp/sysctl.mmcmd.out
This entry should also look the same on cw3

```
--------------------------------------------------------------------------------
#acl#
# These are the users that can issue multinode mmfs commands through sysctl:
_PRINCIPAL root.admin@NWTEST.EMSL.PNL.GOV
_PRINCIPAL root@NWTEST.EMSL.PNL.GOV
_PRINCIPAL rcmd@NWTEST.EMSL.PNL.GOV
--------------------------------------------------------------------------------
```

dsh -av "cat /etc/sysctl.acl" | dshbak -c > /tmp/sysctl.acl.out
This entry should also look the same on CW1

```
--------------------------------------------------------------------------------
#acl#
# This sample acl file contains commented out lines for a principal
# and an acl file.
#_PRINCIPAL root.admin@HPSSL.KGN.IBM.COM
#_ACL_FILE /etc/mycmd.sysctl.acl
_PRINCIPAL root@NWTEST.EMSL.PNL.GOV
_PRINCIPAL root.admin@NWTEST.EMSL.PNL.GOV
_PRINCIPAL rcmd@NWTEST.EMSL.PNL.GOV

--------------------------------------------------------------------------------
```

- **dsh -av "sysctl svcrestart"**
- "**sysctl svcrestart**" on cw3 also
- Test the sysctl authorization by:
  k4list
  sysctl whoami
  sysctl sysctl_vsdcheck
  vsdsklst -a
  vsdsklst -n <pick-a-node-number>
- **dsh -av "sysctl mmremote single"**, you'll want to see an OK at each node response as follows:
- [][cw3][/]>**dsh -av "sysctl mmremote single"**
  t01n01.nwtest.emsl.pnl.gov: ok
  t01n03.nwtest.emsl.pnl.gov: ok
  t01n04.nwtest.emsl.pnl.gov: ok
  t01n05.nwtest.emsl.pnl.gov: ok

```
t01n06.nwtest.emsl.pnl.gov: ok
t01n07.nwtest.emsl.pnl.gov: ok
t01n08.nwtest.emsl.pnl.gov: ok
t01n09.nwtest.emsl.pnl.gov: ok
```

- If there's an authorization problem:
  tail -4 /etc/sysctl.conf, make sure that the following include files are there:

```
# Include VSD sysctl_vsd.cmd commands
include /usr/lpp/csd/sysctl/sysctl_vsd.cmds
# Include mmfs sysctl commands
include /usr/lpp/mmfs/bin/mmcmdsysctl
```

```
dsh -av "stopsrc -s sysctld"
dsh -av "startsrc -s sysctld"
```

2) Steps in creating RVSD's

Now that we have all the software in place and acl's have been configured and tested it's time to create the Virtual Shared Disks and the Recoverable Virtual Shared Disks. The RVSD's are a totally separate activity from GPFS, but GPFS is completely dependent on having properly configured RVSD's. There are several steps to creating the RVSD's:

- "vsdnode" determines which nodes are in the group, both servers and clients.
- Create a "dummy" VSD on any node just to get some subsystems running, this is later removed.
- Determine the "shortest path" between a disk server node and it's SSA disks. We can view this using "diag" and also IBM's tool "maymap". We will use twintailing to improve availability of the SSA disks. Twintailing allows us to completely loose a disk server node but automatically shift the disks over to the designated node for serving.
- Using the list created, run "createvsd" with appropriate parameters for our system.

3) Create the RVSD's (VSD's)

- We enable the VSD's on all 8 nodes (just to make sure), allowing 2 Buddy Buffers on all Client Nodes and 8 Buddy Buffers on all Server Nodes (2 for every disk spindle). The Server Nodes operate the SSA Disk Drives and also run jobs, so pinning unnecessary memory is a bad thing.

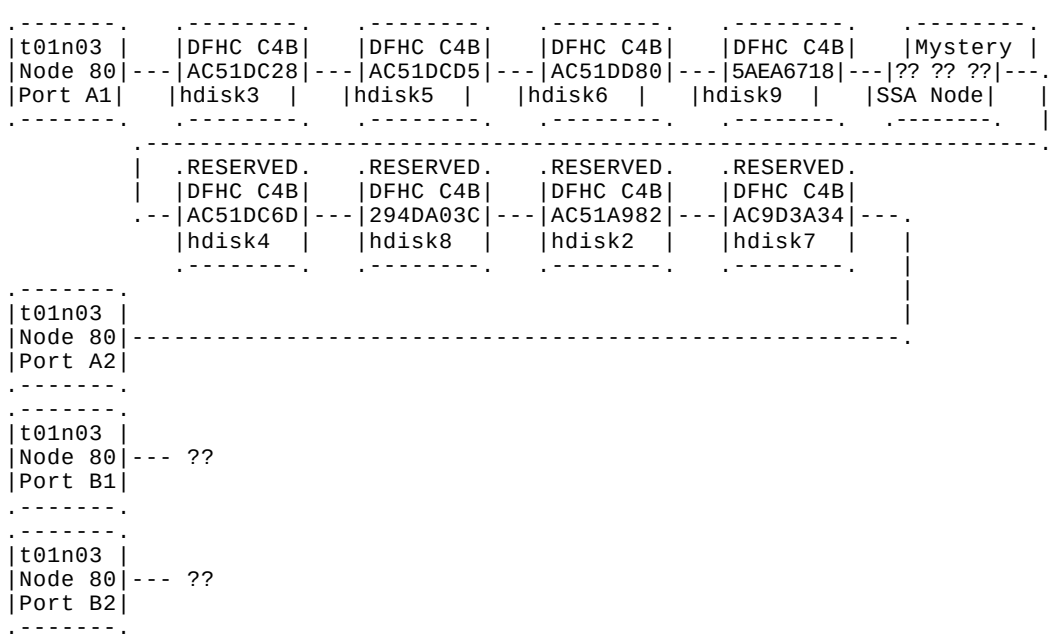
```
/usr/lpp/csd/bin/vsdnode \ 1 5 6 7 8 9 css0 64 256 256 48 4096
262144 2 61440
/usr/lpp/csd/bin/vsdnode \ 3 4 css0 64 256 256 48 4096 262144
8 61440
```

- To verify nodes were properly set up, run "**vsdatalst -n**", a sample follows:

```
        VSD Node Information
                                                Initial Maximum     VSD      rw      Buddy Buffer
  node                  VSD      IP packet   cache   cache request request minimum maximum size: #
number host_name       adapter    size    buffers buffers   count   count    size    size maxbufs
------ --------------- -------- --------- ------- ------- ------- ------- ------- ------- -------
     1 t01n01.nwtest.e css0         61440     256     256     256      48    4096  262144       2
     3 t01n03.nwtest.e css0         61440     256     256     256      48    4096  262144       8
     4 t01n04.nwtest.e css0         61440     256     256     256      48    4096  262144       8
     5 t01n05.nwtest.e css0         61440     256     256     256      48    4096  262144       2
     6 t01n06.nwtest.e css0         61440     256     256     256      48    4096  262144       2
     7 t01n07.nwtest.e css0         61440     256     256     256      48    4096  262144       2
     8 t01n08.nwtest.e css0         61440     256     256     256      48    4096  262144       2
     9 t01n09.nwtest.e css0         61440     256     256     256      48    4096  262144       2
```

- Create a "dummy" VSD on node 1 (t01n01) to activate some subsystems, we will remove it later.
  From t01n01 "**/usr/lpp/csd/bin/createvsd -n 1 -g rootvg -s 8 -c 1 -T 8 -o nocache**"
  From t01n01 "**vsdatalst -v**" to view the new "dummy" vsd and confirm it's name is "vsd1n1".
  From cw3 "**dsh -av /usr/lpp/csd/bin/cfgvsd vsd1n1**"
  From cw3 "**dsh -av /usr/lpp/csd/bin/startvsd vsd1n1**"
  From t01n01 "**/usr/lpp/csd/bin/lsvsd -l**", make sure "ACT" appears in "State" column.
  From t01n01 "**vsdvts vsd1n1**" to test the dummy vsd, "Continue? YES"
  To cleanup this dummy vsd:
  From cw3 "**dsh -av suspendvsd vsd1n1**"
  From cw3 "**dsh -av stopvsd vsd1n1**"
  From cw3 "**dsh -av ucfgvsd vsd1n1**"
  From cw3 "**removevsd -v vsd1n1 -f**" to force removal of this dummy vsd.
  From cw3 "**dsh -av lsvsd -l**", you should get errors from every node.

- Create the VSD's for the 8 disks in this gpfs filesystem. We twintail SSA Disks for availability reasons. T more easily understand how this is done I placed a tool calld "maymap" into /systems/gpfstest/maymap. Running "maymap -h" gives the following:



The "Mystery" box is actually the twintail node these disks are shared by, in this case t01n04 or node 4. Notice there are not any connections to the B Ports on the SSA Adapter Card. There are options with "maymap" that allow output other than "boxes", running "maymap -lph" gives the following:

```
************************************************************************
Resrc ID Serial No UID      : A1 A2 B1 B2 HWI Status      Model    Code Lvl   Reserved  pdisk    hdisk
0301000E 68B27216  AC51DC28 : 00 08 FF FF (0=ST_Good)     DFHC C4B RAMST077             pdisk1   hdisk3
0301000B 68A88263  AC51DCD5 : 01 07 FF FF (0=ST_Good)     DFHC C4B RAMST077             pdisk3   hdisk5
0301000F 68B27339  AC51DD80 : 02 06 FF FF (0=ST_Good)     DFHC C4B RAMST077             pdisk4   hdisk6
03010009 68191664  5AEA6718 : 03 05 FF FF (0=ST_Good)     DFHC C4B RAMST077             pdisk7   hdisk9
0301000D 68B23414  AC51DC6D : 05 03 FF FF (1=ST_Failed)   DFHC C4B RAMST077   Reserved  pdisk2   hdisk4
03010008 680E78F5  294DA03C : 06 02 FF FF (1=ST_Failed)   DFHC C4B RAMSC095   Reserved  pdisk6   hdisk8
0301000C 68204021  AC51A982 : 07 01 FF FF (1=ST_Failed)   DFHC C4B RAMST077   Reserved  pdisk0   hdisk2
0301000A 68B19319  AC9D3A34 : 08 00 FF FF (1=ST_Failed)   DFHC C4B RAMST077   Reserved  pdisk5   hdisk7

************************************************************************
```

The column with the "reserved" refers to SSA Disks assigned to this serving node's twintail partner node. If you don't have maymap, enter "diag", "Task Selection(Diagnostics, Advanced Diagnostics, Service Aids, etc.)", "SSA Service Aids", "Link Verification", "et0101:ssa0 00-03 SSA Adapter" which will display the following:

Physical	Serial#	Adapter Port A1	A2	B1	B2	Status
t01n03:pdisk1	AC51DC28	0	8			Good
t01n03:pdisk3	AC51DCD5	1	7			Good
t01n03:pdisk4	AC51DD80	2	6			Good
t01n03:pdisk7	5AEA6718	3	5			Good
*****		4	4			
t01n03:pdisk2	AC51DC6D	5	3			Reserved
t01n03:pdisk6	294DA03C	6	2			Reserved
t01n03:pdisk0	AC51A982	7	1			Reserved
t01n03:pdisk5	AC9D3A34	8	0			Reserved

This procedure requires that you determine the relationship between "pdisk" and "hdisk" since hdisk is required to create the vsd's and I use the pdisk the give a volume name to the disk which helps in problem determination. Step back a couple screen on the diag till you see "Configuration Verification" and follow that path, it will display the pdisk to hdisk relationship as follows:

```
t01n03:pdisk0                   AC51A982  4GB SSA C Physical Disk Drive
t01n03:pdisk1                   AC51DC28  4GB SSA C Physical Disk Drive
t01n03:pdisk2                   AC51DC6D  4GB SSA C Physical Disk Drive
t01n03:pdisk3                   AC51DCD5  4GB SSA C Physical Disk Drive
t01n03:pdisk4                   AC51DD80  4GB SSA C Physical Disk Drive
t01n03:pdisk5                   AC9D3A34  4GB SSA C Physical Disk Drive
t01n03:pdisk6                   294DA03C  4GB SSA C Physical Disk Drive
t01n03:pdisk7                   5AEA6718  4GB SSA C Physical Disk Drive
t01n03:hdisk2                   AC51A982  SSA Logical Disk Drive
t01n03:hdisk3                   AC51DC28  SSA Logical Disk Drive
t01n03:hdisk4                   AC51DC6D  SSA Logical Disk Drive
t01n03:hdisk5                   AC51DCD5  SSA Logical Disk Drive
t01n03:hdisk6                   AC51DD80  SSA Logical Disk Drive
t01n03:hdisk7                   AC9D3A34  SSA Logical Disk Drive
t01n03:hdisk8                   294DA03C  SSA Logical Disk Drive
t01n03:hdisk9                   5AEA6718  SSA Logical Disk Drive
```

After all this is determined, I created a file that that was executable.

The following is the contents of the createvsd.gpfs_servers.sh file:

```
createvsd -n 3/4:hdisk3/ -s 2148 -g pd1 -T 4 -o nocache
createvsd -n 3/4:hdisk5/ -s 2148 -g pd3 -T 4 -o nocache
createvsd -n 3/4:hdisk6/ -s 2148 -g pd4 -T 4 -o nocache
createvsd -n 3/4:hdisk9/ -s 2148 -g pd7 -T 4 -o nocache
#
createvsd -n 4/3:hdisk2/ -s 2148 -g pd2 -T 4 -o nocache
createvsd -n 4/3:hdisk4/ -s 2148 -g pd6 -T 4 -o nocache
createvsd -n 4/3:hdisk7/ -s 2148 -g pd0 -T 4 -o nocache
createvsd -n 4/3:hdisk8/ -s 2148 -g pd5 -T 4 -o nocache
```

Every line entry will create output similar to the following:

```
********************************************************************************
+ createvsd -n 1/49:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd: calls Getopts.
createvsd: parsing node_list.
createvsd: call vsd_rollback.
It took about 1 seconds in vsd_rollback.
createvsd: creates task tables.
createvsd: calls sysctl_mkvglv.
It took about 10 seconds in mkvglv.
createvsd: calls sysctl_varyoffvg 15vg on the primary node cef1.nwmpp1.emsl.pnl.gov
createvsd: calls sysctl_importvg 15vg on the backup node cef4.nwmpp1.emsl.pnl.gov  with
00006186d171b9ff
It took about 28 seconds in importvg.
createvsd: calls sysctl_varyonvg 15vg on the primary node cef1.nwmpp1.emsl.pnl.gov
createvsd: calls vsdvg.
OK:0:vsdvg -g 15vgn1b49 15vg 1 49
It took about 3 seconds in vsdvg.
createvsd: calls defvsd.
OK:0:defvsd 15lv4n1 15vgn1b49 15v4n1 nocache
It took about 3 seconds in defvsd.
********************************************************************************
```

Configure the newly created VSD's by running "**cfgvsd -a**" on cw3, then "**lsvsd -l**" from t01n01 to verify they are all there and configured. We will not start them or make them active, the GPFS/MMFS code will do that for us.

4) Creating GPFS File Systems

- Configure and start the GPFS File System; login to t01n01 and run:
  "**k4init root.admin**"
  "**dsh -av -a startsrc -s mmfs**"
  "**/usr/lpp/mmfs/bin/mmconfig -a -A -m 4M -M 200 -p 20M**" will configure all nodes with mallocsize of 4M, MaxFilesToCache=200 and pagepool=20M, while the subsystem will automatically start-up. To verify, cat /var/mmfs/etc/mmfs.cfg and view the options.

- Create File Descriptions that MMFS needs, call it /var/mmfs/scratchdesc for /gpfs. Place this file in t01n01 which is where the mmfs commands will run from. This file consists of all the VSD names found in the first column of the command "vsdatalst -v" that are to serve a particular gpfs filesystem. We use one filesystems, /gpfs with 4 2.2 GB Disks. An example of the first couple of VSD names for /gpfs follows (just use the first column):

VSD Table

VSD name	logical volume	Global Volume Group	minor#	option	size_in_MB
vsd1n3	lvvsd1n3	pd1n3b4	1	nocache	4200
vsd2n3	lvvsd2n3	pd3n3b4	2	nocache	4200
vsd3n3	lvvsd3n3	pd4n3b4	3	nocache	4200
vsd4n3	lvvsd4n3	pd7n3b4	4	nocache	4200
vsd5n4	lvvsd5n4	pd2n4b3	5	nocache	4200
vsd6n4	lvvsd6n4	pd6n4b3	6	nocache	4200
vsd7n4	lvvsd7n4	pd0n4b3	7	nocache	4200
vsd8n4	lvvsd8n4	pd5n4b3	8	nocache	4200

- To actually create the /gpfs filesystem, run: "**/usr/lpp/mmfs/bin/mmcrfs /gpfs /dev/scratchfs -F /var/mmfs/scratchdesc -A no -B 256K -i 4K -I 16K -n 32 -N 128K -v no -R 1 -M 1 -r 1 -m 1 -Q no -s roundRobin**"
- After completion, run "**mmlsfs /dev/gpfs**" from t01n01 to get:

```
flag value          description
---- -------------- -----------------------------------------------------
 -s  roundRobin     Stripe method
 -f  8192           Minimum fragment size in bytes
 -i  4096           Inode size in bytes
 -I  16384          Indirect block size in bytes
 -m  1              Default number of metadata replicas
 -M  1              Maximum number of metadata replicas
 -r  1              Default number of data replicas
 -R  1              Maximum number of data replicas
 -a  268800         Estimated average file size
 -n  32             Estimated number of nodes that will mount file system
 -B  262144         Block size
 -Q  none           Quotas enforced
 -F  131584         Maximum number of inodes
 -V  2              File system version. Highest supported version: 2
 -d  vsd1n3;vsd2n3;vsd3n3;vsd4n3;vsd5n4;vsd6n4;vsd7n4;vsd8n4  Disks in file system
 -C  1              Configuration identifier
```

- Mount the new filesystems "**dsh -av mount /gpfs**"
- We use gpfs for scratch space for our users. To do this requires that user directories be created on the files space, one directory for every user. Run "**/usr/local/sbin/add_all_gpfs_users**" from t01n01.

d) Installation Basics Thin/High Node SP NWecs1

1) Accomplish the same tasks you did above for software, sysctl's and switch tuning. Then run the following:
   - **dsh -av "sysctl svcrestart"**
   - "**sysctl svcrestart**" on cw4 also
   - Test the sysctl authorization by:
     k4list
     sysctl whoami
     sysctl sysctl_vsdcheck
     vsdsklst -a
     vsdsklst -n <pick-a-node-number>
     dsh -av "sysctl mmremote single", you'll want to see an OK at each node response as follows:

```
[][cw3][/]>dsh -av "sysctl mmremote single"
x01n01.ecs1.emsl.pnl.gov: ok
x01n02.ecs1.emsl.pnl.gov: ok
x01n03.ecs1.emsl.pnl.gov: ok
x01n04.ecs1.emsl.pnl.gov: ok
x01n05.ecs1.emsl.pnl.gov: ok
x01n06.ecs1.emsl.pnl.gov: ok
x01n07.ecs1.emsl.pnl.gov: ok
x01n08.ecs1.emsl.pnl.gov: ok
x01n09.ecs1.emsl.pnl.gov: ok
x01n10.ecs1.emsl.pnl.gov: ok
x01n11.ecs1.emsl.pnl.gov: ok
x01n12.ecs1.emsl.pnl.gov: ok
x01n13.ecs1.emsl.pnl.gov: ok
x01n14.ecs1.emsl.pnl.gov: ok
x01n15.ecs1.emsl.pnl.gov: ok
x01n16.ecs1.emsl.pnl.gov: ok
x02n01.ecs1.emsl.pnl.gov: ok
x02n05.ecs1.emsl.pnl.gov: ok
x02n09.ecs1.emsl.pnl.gov: ok
x02n13.ecs1.emsl.pnl.gov: ok
```

- If there's an authorization problem:
  tail -4 /etc/sysctl.conf, make sure that the following include files are there:
  # Include VSD sysctl_vsd.cmd commands
  include /usr/lpp/csd/sysctl/sysctl_vsd.cmds
  # Include mmfs sysctl commands
  include /usr/lpp/mmfs/bin/mmcmdsysctl
  dsh -av "stopsrc -s sysctld"
  dsh -av "startsrc -s sysctld"

2) Steps in creating RVSD's

Now that we have all the software in place and acl's have been configured and tested it's time to create the Virtual Shared Disks and the Recoverable Virtual Shared Disks. The RVSD's are a totally separate activity from GPFS, but GPFS is completely dependent on having properly configured RVSD's. There are several steps to creating the RVSD's:

- "**vsdnode**" determines which nodes are in the group, both servers and clients.
- Create a "dummy" VSD on any node just to get some subsystems running, this is later removed.
- Determine the "shortest path" between a disk server node and it's SSA disks. We can view this using "diag" and also IBM's tool "maymap". We will use twintailing to improve availability of the SSA disks. Twintailing allows us to completely loose a disk server node but automatically shift the disks over to the designated node for serving.
- Using the list created, run "createvsd" with appropriate parameters for our system.

3) Create the RVSD's (VSD's)

- We enable the VSD's on all 20 nodes (just to make sure), allowing 2 Buddy Buffers on all Client Nodes and 6 Buddy Buffers on all Server Nodes. The Server Nodes operate the SSA Disk Drives and also run jobs, so pinning unnecessary memory is a bad thing.

  /usr/lpp/csd/bin/vsdnode \ 1 4 5 6 7 8 9 10 11 12 13 14 15 16 17 21 25 29 css0 256 256 256 48 4096 262144 2 61440

  /usr/lpp/csd/bin/vsdnode \ 2 3 css0 256 256 256 48 4096 262144 6 61440

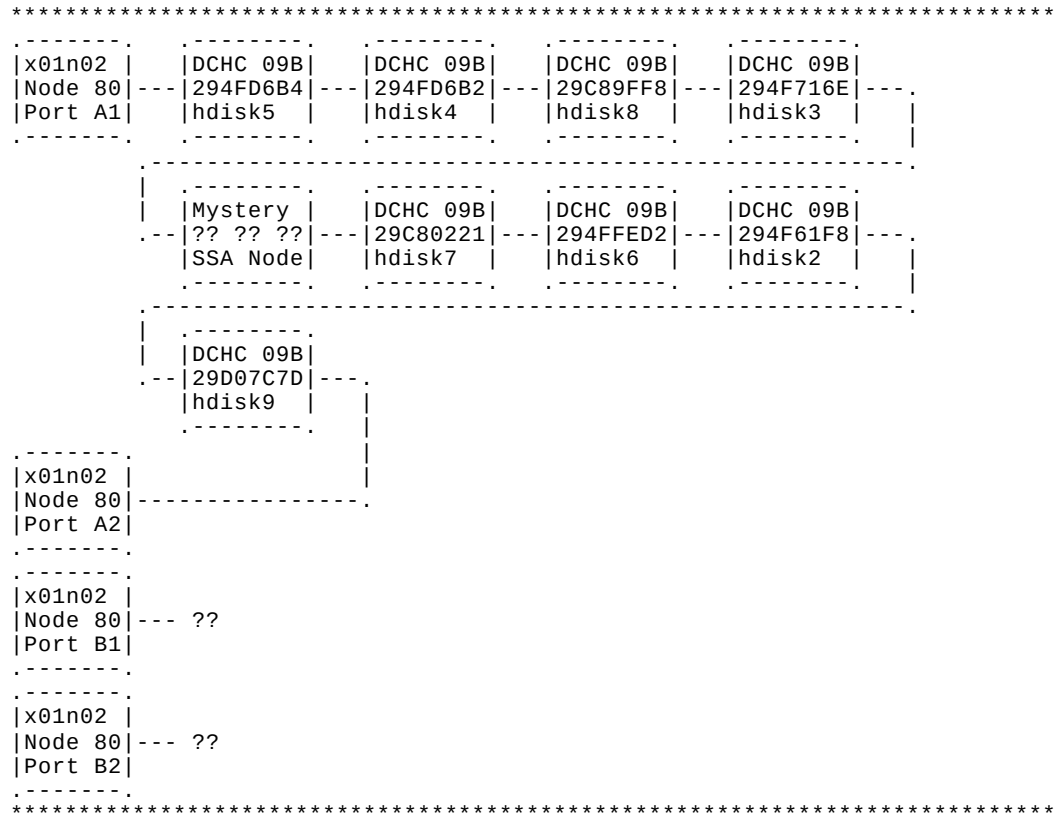
To verify nodes were properly set up, run "**vsdatalst -n**", a sample follows:

```
       VSD Node Information
                                                Initial Maximum     VSD      rw      Buddy Buffer
  node                    VSD      IP packet   cache   cache request request minimum maximum size: #
number host_name        adapter    size    buffers buffers   count   count    size    size maxbufs
------ ---------------- -------- --------- ------- ------- ------- ------- ------- ------- -------
     1 x01n01.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
     2 x01n02.ecs1.ems css0         61440     256     256     256      48    4096  262144       6
     3 x01n03.ecs1.ems css0         61440     256     256     256      48    4096  262144       6
     4 x01n04.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
     5 x01n05.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
     6 x01n06.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
     7 x01n07.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
     8 x01n08.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
     9 x01n09.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    10 x01n10.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    11 x01n11.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    12 x01n12.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    13 x01n13.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    14 x01n14.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    15 x01n15.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    16 x01n16.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    17 x02n01.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    21 x02n05.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    25 x02n09.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
    29 x02n13.ecs1.ems css0         61440     256     256     256      48    4096  262144       2
```

- Create a "dummy" VSD on node 1 (x01n01) to activate some subsystems, we will remove it later.

  From x01n01 "**/usr/lpp/csd/bin/createvsd -n 1 -g rootvg -s 8 -c 1 -T 16 -o nocache**"
  From x01n01 "**vsdatalst -v**" to view the new "dummy" vsd and confirm it's name is "vsd1n1".
  From cw4 "**dsh -av /usr/lpp/csd/bin/cfgvsd vsd1n1**"
  From cw4 "**dsh -av /usr/lpp/csd/bin/startvsd vsd1n1**"
  From x01n01 "**/usr/lpp/csd/bin/lsvsd -l**", make sure "ACT" appears in "State" column.
  From x01n01 "**vsdvts vsd1n1**" to test the dummy vsd, "Continue? YES"
  To cleanup this dummy vsd:
  From cw4 "**dsh -av suspendvsd vsd1n1**"
  From cw4 "**dsh -av stopvsd vsd1n1**"
  From cw4 "**dsh -av ucfgvsd vsd1n1**"
  From cw4 "**removevsd -v vsd1n1 -f**" to force removal of this dummy vsd.
  From cw4 "**dsh -av lsvsd -l**", you should get errors from every node.

- Create the VSD's for the 8 disks in this gpfs filesystem. We twintail SSA Disks for availability reasons. T more easily understand how this is done I placed a tool calld "maymap" into /systems/gpfstest/maymap. Running "maymap -h" gives the following:



The "Mystery" box is actually the twintail node these disks are shared by, in this case x01n03 or node 3. Notice there are not any connections to the B Ports on the SSA Adapter Card. There are options with "maymap" that allow output other than "boxes", running "maymap -lph" gives the following:

```
*******************************************************************************
Resrc ID Serial No UID      : A1 A2 B1 B2 HWI Status      Model    Code Lvl Reserved  pdisk    hdisk
03000007 680BD519  294FD6B4 : 00 08 FF FF (0=ST_Good)     DCHC 09B RAMSC095          pdisk3   hdisk5
03000006 680BD4A5  294FD6B2 : 01 07 FF FF (0=ST_Good)     DCHC 09B RAMSC095          pdisk2   hdisk4
03000005 6810D8FB  29C89FF8 : 02 06 FF FF (0=ST_Good)     DCHC 09B RAMSC095          pdisk6   hdisk8
03000004 680C4616  294F716E : 03 05 FF FF (0=ST_Good)     DCHC 09B RAMSC095          pdisk1   hdisk3
03000003 6809927E  29C80221 : 05 03 FF FF (0=ST_Good)     DCHC 09B RAMSC095          pdisk5   hdisk7
03000002 680BCA6C  294FFED2 : 06 02 FF FF (0=ST_Good)     DCHC 09B RAMSC095          pdisk4   hdisk6
03000001 680B8E30  294F61F8 : 07 01 FF FF (0=ST_Good)     DCHC 09B RAMSC095          pdisk0   hdisk2
03000000 6813B783  29D07C7D : 08 00 FF FF (0=ST_Good)     DCHC 09B CUSMA903          pdisk7   hdisk9

*******************************************************************************
```

The column with the "reserved" refers to SSA Disks assigned to this serving node's twintail partner node. If you don't have maymap, enter "diag", "Task Selection(Diagnostics, Advanced Diagnostics, Service Aids, etc.)", "SSA Service Aids", "Link Verification", "x01n02:ssa0 00-03 SSA Adapter" which will display the following:

```
  Physical                          Serial#     Adapter Port
                                                A1  A2  B1  B2      Status

  x01n02:pdisk3                     294FD6B4    0   8               Good
  x01n02:pdisk2                     294FD6B2    1   7               Good
  x01n02:pdisk6                     29C89FF8    2   6               Good
  x01n02:pdisk1                     294F716E    3   5               Good
  x01n03:ssa0:A                                 4   4
  x01n02:pdisk5                     29C80221    5   3               Good
  x01n02:pdisk4                     294FFED2    6   2               Good
  x01n02:pdisk0                     294F61F8    7   1               Good
  x01n02:pdisk7                     29D07C7D    8   0               Good
```

This procedure requires that you determine the relationship between "pdisk" and "hdisk" since hdisk is required to create the vsd's and I use the pdisk the give a volume name to the disk which helps in problem determination. Step back a couple screen on the diag till you see "Configuration Verification" and follow that path, it will display the pdisk to hdisk relationship as follows:

```
x01n02:pdisk0                   294F61F8  9.1GB SSA C Physical Disk Drive
x01n02:pdisk1                   294F716E  9.1GB SSA C Physical Disk Drive
x01n02:pdisk2                   294FD6B2  9.1GB SSA C Physical Disk Drive
x01n02:pdisk3                   294FD6B4  9.1GB SSA C Physical Disk Drive
x01n02:pdisk4                   294FFED2  9.1GB SSA C Physical Disk Drive
x01n02:pdisk5                   29C80221  9.1GB SSA C Physical Disk Drive
x01n02:pdisk6                   29C89FF8  9.1GB SSA C Physical Disk Drive
x01n02:pdisk7                   29D07C7D  9.1GB SSA C Physical Disk Drive
x01n02:hdisk2                   294F61F8  SSA Logical Disk Drive
x01n02:hdisk3                   294F716E  SSA Logical Disk Drive
x01n02:hdisk4                   294FD6B2  SSA Logical Disk Drive
x01n02:hdisk5                   294FD6B4  SSA Logical Disk Drive
x01n02:hdisk6                   294FFED2  SSA Logical Disk Drive
x01n02:hdisk7                   29C80221  SSA Logical Disk Drive
x01n02:hdisk8                   29C89FF8  SSA Logical Disk Drive
x01n02:hdisk9                   29D07C7D  SSA Logical Disk Drive
```

After all this is determined, I created a file that I will run. These are 9.1GB SSA Disks so the proper physical partition and disk size must be included in the options on createvsd as follows.

The following is the contents of the createvsd.gpfs_servers.sh file should be executable:

```
createvsd -n 2/3:hdisk5/ -s 8672 -g n2hd5 -T 16 -o nocache
createvsd -n 2/3:hdisk4/ -s 8672 -g n2hd4 -T 16 -o nocache
createvsd -n 2/3:hdisk8/ -s 8672 -g n2hd8 -T 16 -o nocache
createvsd -n 2/3:hdisk3/ -s 8672 -g n2hd3 -T 16 -o nocache
#
createvsd -n 3/2:hdisk7/ -s 8672 -g n3hd7 -T 16 -o nocache
createvsd -n 3/2:hdisk6/ -s 8672 -g n3hd6 -T 16 -o nocache
createvsd -n 3/2:hdisk2/ -s 8672 -g n3hd2 -T 16 -o nocache
createvsd -n 3/2:hdisk9/ -s 8672 -g n3hd9 -T 16 -o nocache
```

Every line entry will create output similar to the following:

```
********************************************************************************
+ createvsd -n 1/49:hdisk15/ -s 2148 -g 15vg -v 15v -l 15lv -T 4 -o nocache
createvsd: calls Getopts.
createvsd: parsing node_list.
createvsd: call vsd_rollback.
It took about 1 seconds in vsd_rollback.
createvsd: creates task tables.
createvsd: calls sysctl_mkvglv.
It took about 10 seconds in mkvglv.
createvsd: calls sysctl_varyoffvg 15vg on the primary node
cef1.nwmpp1.emsl.pnl.gov
createvsd: calls sysctl_importvg 15vg on the backup node cef4.nwmpp1.emsl.pnl.gov
with 00006186d171b9ff
It took about 28 seconds in importvg.
createvsd: calls sysctl_varyonvg 15vg on the primary node
cef1.nwmpp1.emsl.pnl.gov
createvsd: calls vsdvg.
OK:0:vsdvg -g 15vgn1b49 15vg 1 49
It took about 3 seconds in vsdvg.
createvsd: calls defvsd.
OK:0:defvsd 15lv4n1 15vgn1b49 15v4n1 nocache
It took about 3 seconds in defvsd.
********************************************************************************
```

Configure the newly created VSD's by running "**cfgvsd -a**" on cw4, then "**lsvsd -l**" from x01n01 to verify they are all there and configured. We will not start them or make them active, the GPFS/MMFS code will do that for us, but once they are started they will look like this:

```
minor  state server lv_major lv_minor vsd-name    option    size(MB)
 1      ACT    2      39       1      vsd1n2      nocache       8192
 2      ACT    2      40       1      vsd2n2      nocache       8192
 3      ACT    2      41       1      vsd3n2      nocache       8192
 4      ACT    3       0       0      vsd4n3      nocache       8192
 5      ACT    3       0       0      vsd5n3      nocache       8192
 6      ACT    3       0       0      vsd6n3      nocache       8192
 7      ACT    3       0       0      vsd7n3      nocache       8192
 8      ACT    2      46       1      vsd8n2      nocache       8192
```

4) Creating GPFS File Systems

- Configure and start the GPFS File System; login to x01n01 and run:
  "**k4init root.admin**"
  "**dsh -av -a startsrc -s mmfs**"
  "**/usr/lpp/mmfs/bin/mmconfig -a -A -m 16M -M 1000 -p 64M**" will configure all nodes with mallocsize of 16M, MaxFilesToCache=1000 and pagepool=64M, while the subsystem will automatically start-up. To verify, cat /var/mmfs/etc/mmfs.cfg and view the options.

- Create File Descriptions that MMFS needs, call it /var/mmfs/scratchdesc for /gpfs. Place this file in x01n01 which is where the mmfs commands will run from. This file consists of all the VSD names found in the first column of the command "vsdatalst -v" that are to serve a particular gpfs filesystem. We use one filesystems, /gpfs with 8 9.1 GB Disks. The output from the vsdatalst -v command appears next (just use the first column):

```
VSD name                        logical volume  Global Volume Group             minor# option
size_in_MB
------------------------------- --------------- ------------------------------ ------ --------- ----
vsd1n2                            lvvsd1n2        n2hd3n2b3                       1 nocache    8192
vsd2n2                          lvvsd2n2        n2hd4n2b3                       2 nocache    8192
vsd3n2                          lvvsd3n2        n2hd5n2b3                       3 nocache    8192
vsd4n3                          lvvsd4n3        n3hd6n3b2                       4 nocache    8192
vsd5n3                          lvvsd5n3        n3hd7n3b2                       5 nocache    8192
vsd6n3                            lvvsd6n3        n3hd8n3b2                       6 nocache    8192
vsd7n3                          lvvsd7n3        n3hd9n3b2                       7 nocache    8192
vsd8n2                          lvvsd8n2        n2hd2n2b3                       8 nocache    8192
```

- To actually create the /gpfs filesystem, run: "**/usr/lpp/mmfs/bin/mmcrfs /gpfs /dev/scratchfs -F /var/mmfs/scratchdesc -A no -B 256K -i 4K -I 16K -n 32 -N 128K -v no -R 1 -M 1 -r 1 -m 1 -Q no -s roundRobin -v no**"
- After completion, run "**mmlsfs /dev/gpfs**" from x01n01 to get:

```
flag value          description
---- -------------- -----------------------------------------------------
 -s  roundRobin     Stripe method
 -f  8192           Minimum fragment size in bytes
 -i  512            Inode size in bytes
 -I  16384          Indirect block size in bytes
 -m  1              Default number of metadata replicas
 -M  1              Maximum number of metadata replicas
 -r  1              Default number of data replicas
 -R  1              Maximum number of data replicas
 -a  1048576        Estimated average file size
 -n  32             Estimated number of nodes that will mount file system
 -B  262144         Block size
 -Q  none           Quotas enforced
 -F  69632          Maximum number of inodes
 -V  2              File system version. Highest supported version: 2
 -d  vsd1n2;vsd2n2;vsd3n2;vsd4n3;vsd5n3;vsd6n3;vsd7n3;vsd8n2  Disks in file
system
 -C  1              Configuration identifier
```

- Mount the new filesystems "dsh -av mount /gpfs"
- We use gpfs for scratch space for our users. To do this requires that user directories be created on the files space, one directory for every user. Run "**/usr/local/sbin/add_all_gpfs_users**" from x01n01.
- See Previous section for starting, mounting and recovery.

5) Good things to know
   - **smitty delfs**
   - **dsh -w cw1 -a "stopsrc -s sysctld"**
   - **dsh -w cw1 -a "startsrc -s sysctld"**
   - **"kinit root.admin" (on any CWS or node you're running from)**
   - **"mmfsadm shutdown" (to properly stop mmfs)**
   - **dsh -av "startsrc -s mmfs"**
   - **suspendvsd –a**
   - **stopvsd –a**
   - **ucfgvsd –a**
   - **cfgvsd –a**
   - **startvsd –a**
   - **smitty delete_vsd (careful with this one)**

- **mmfsadm cleanup**
- **lsvsd –l**
- **vsdatalst -v (-n, -g)**
- **mmdf /dev/gpfs**
- **mmlsfs /dev/gpfs**
- **lssrc -ls hags**
- **phoenix.snap**
- **mmfsadm dump all**
- **mmfsadm dump cfgmgr**
- **restripefs**
- **mmfsck**
- **mmaddnode**
- **/usr/sbin/rsct/bin/hagsgr -s hags**
- **/usr/sbin/rsct/bin/hagsvote -ls hags**
- Run "**lsscr -ls hags**" from a switch node to view the following:

```
Subsystem         Group            PID     Status
 hags             hags             17850   active
 5 locally-connected clients.  Their PIDs:
 643824 11710 19614 21826 10436
 HA Group Services domain information:
 Domain established by node 0.
 Number of groups known locally: 5
                     Number of   Number of local
 Group name          providers   providers/subscribers
 Gpfs.1                  506           1           0
 GpfsRec.1               506           1           0
 cssMembership           512           1           2
 ha_em_peers             513           1           0
 ha.vsd                  506           1           0
```

We have 512 switch nodes, add 1 for the CWs to get 513 peers. Only 506 nodes are currently running the mmfs codeby design. Knowing this is very handy for problem determination.

6) Recovering from a major goof up when remove commands don't work:
   - Manually remove all VSD Logical Volumes, vary them off and export them out of here.
   - "**smitty delete_vsd**" and wipe out everything.
   - "**SDRListFiles**" and look for mmsdrfs, the default name for gpfs filess in the SDR.
   - "**SDRRetrieveFile mmsdrfs /tmp/edit**"
   - "**vi /tmp/edit**" and remove entries from /gpfs on down (leave the node list).
   - "**SDRReplaceFile /tmp/edit mmsdrfs**"
   - Now you are ready to run the "vsdnode" commands listed in an earlier step to create the VSD's.
   - Monitor the accuracy of the selected parameters by running "**/usr/lpp/csd/bin/statvsd**" on all nodes. Pay particular attention to "requests queued waiting for buddy buffer". We set 2 buddy buffers for every client node and 64 for every server node, that choice might need to be tuned depending on results. Running "updatevsdnode" with appropriate parameters

will fix this problem. Remember to stop the mmfs processes before altering the VSD's.

- A good place to view VSD status is "/var/adm/csd/vsd.log", a good sign is when "/usr/lpp/csd/bin/vsd.FENCE3: Finished" shows up on the output.
- The best place to view mmfs status is "/var/adm/ras/mmfs.log.xxxx", a good sign is the following:

```
/usr/lpp/mmfs/bin/runmmfs starting
mmfsd initializing.
mmfsd ready
mounting /gpfs1
```

7) Workaround for startup-delay
   - We get 2 definite delays with our 512 node system on any node that we start mmfs on. Our current workaround is: Edit /var/mmfs/etc/mmfs.cfg by adding "trace cleanup 1" after "trace all 0" in the file.
   - Starting mmfs via "startsrc -e "NSORDER=bind,local" -s mmfs"
   - This allows starting mmfs and mounting the filesystems in around 30 seconds as opposed to around 7 minutes.

8) Replacing a Disk
   - Hard drives go bad, its just so. Since we use our GPFS as scratch only, we do not need to backup and restore any data.
   - If the disk is still kind of running, "suspend" the vsd first if possible, then run "mmdeldisk /gpfs1 9v1n209 -p. The "-p" means permanently damaged, well we want to replace it anyway. We cannot run "mmrpldisk" as we do not have any spare slots to add an additional hard drive to.
   - On this 512 node system, mmdeldisk takes 3hrs 45 minutes to fail then 1 hr and 40 minutes to recover. Expect a successful mmdeldisk to take most of the day, it has to edit the SDR and every /etc/filesystem entry on every node to remove any mention of the errant vsd you want to delete.
   - Once the VSD is gone (verify with a lsvsd -l), go to the primary node serving that bad disk, remove any and all mention of that disk (smitty devices, then SSA) and do the same for the twintailed node. Now physically remove the disk, replace with a new one, the boot up will do the cfgmgr and the new disk will slide into hdisk and pdisk definitions identical to the old one but with a new serial number. Run cfgmgr on the twintailed node to recognize the new disk. Run the appropriate entry from the executable "createvsd" script to define the new SSA disk:

```
createvsd -n 227/211:hdisk3/  -s 2148 -g  3vg -v  3v -l  3lv -T 4
-o nocache
```

   - Verify the name created matches the old VSD name (lsvsd -l). Configure using smitty and make active by running "ha.vsd reset".
   - Now do a mmadddisk for GPFS to take in the new disk.

71. *Configure nfs home directories on CWS*

a) smitty mknfsmnt
b) mount now and add to /etc/filesystems: both
c) mount type: hard
d) /u1 - prophet; /u2 - prophet

72. *Setup ATM gateway nodes* - nwmpp1, nwecs1 only

a) telnet into each node and do the following:
(1) smitty install_selectable_all /systems/lpp/aix432
(2) Select appropriate device software for the architecture type (pci, mca, isa).
(3) shutdown -Fr
(4) netstat -v (Get the atm interface MAC address)
(5) smitty atmle_panel
  (a) Add an ATM LE Client
  (b) Add an Ethernet ATM LE Client
  (c) Info to enter:
    MAC Address
    Auto config yes
    Emulated LAN name (Supplied by network administrator)
(6) **smitty mkinet**
  (a) Add A Standard Ethernet Network Interface
  (b) Info to enter:
  (c) Select en1
  (d) Enter IP address (Supplied by network admin)
  (e) Enter netmask (Supplied by network admin)
(7) **smitty route**
  Add a static route
  Destination type: net
  Destination address: default
  Default GATEWAY Address: (Supplied by network admin)
  Check default route (netstat -rn)
(8) ping default gateway
(9) Set the hostname for the ATM interface
b) Ensure this is not the same as the default 'hostname'

73. *Configure Maui (Latest Available Version)*

a) **cd /systems/src/**
b) ls -ltr maui*
c) **cp -p maui.<latest>.tar /loadl**
d) **su - loadl**
e) **tar -tvf maui**. . . This should list all the files, verify that they'll go into the maui directory.
f) **tar -xvf maui.<latest>.tar**
g) **cd maui; ./setupmaui**
h) Verify that the maui.cfg file is correct
i) Check the /loadl/LoadL_config file to make sure that *SCHEDULER_API = YES.*
j) **llctl -g stop; llctl -g start**
k) Check to see that */loadl/maui/bin* is in loadl's path
l) **Maui**

74. *Configure ADSM clients*

a) On each ADSM client (currently cw1, et1507, et1513, seer)
b) cd /systems/lpp/adsm
c) smitty install_latest
d) install the client software

75. *Configure crontabs*

76. *Install qbank.*
See:http://www.emsl.pnl.gov:2080/docs/mscf/qbankbeta/install/install.html for pre-req's and pre-req's installation hints.

77. *Install / Configure ssh - Site Specific for PNNL*

a) AFS must be available
- /systems/bin/rmdce22.expect
- /systems/bin/mkdce22.expect

b) There is a webpage -

http://www.emsl.pnl.gov/auth/cool/system_procedures/ssh/ssh_admin.html#configure

c) Check that the current @sys is correct by issuing **/usr/afsws/bin/fs sysiname** (if aix 432 then you should see rs_aix43)
d) If not correct then issue **/usr/afsws/bin/fs sysiname <sysname>**
e) If ssh has already been installed, consider saving /krb5/krb5.keytab. Having this should make it so you don't have to regenerate keys (step 2). Additionally, re-installing ssh will give users the following error message when they try connect:
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
WARNING: HOST IDENTIFICATION HAS CHANGED!
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
IT IS POSSIBLE THAT SOMEONE IS DOING SOMETHING NASTY!
Someone could be eavesdropping on you right now (man-in-the-middle attack)! It is also possible that the host key has just been changed. Please contact your system administrator. Add correct host key in /dfs/home/dhazen/.ssh/known_hosts to get rid of this message. Agent forwarding is disabled to avoid attacks by corrupted servers. X11 forwarding is disabled to avoid attacks by corrupted servers.This can be avoided by saving off the host key files in /etc. The drawback is that you then have to do a manual install (not documented here) instead of using the install scripts provided. Users can get rid of this message by removing the host key entry in ~/.ssh/known_hosts.

(1) Configure kerberos **/afs/msrc/wsadmin/share/bin/config_k5-1.0.5** The configuration script may put 'extra' entries in /etc/services and /etc/inetd.conf. Remove any duplicates that may exist for: *klogin kshell eklogin* If you modified /etc/inetd.conf, then do: **refresh -s inetd.conf**

(2) Generate kerberos principal and local keytab entry (do as cell_admin) **/msrc/home/admin/scripts/keytab_admin.sh** Note that this script has been broken so that it generates an incorrect machine name - eg [t01n01][/krb5]> /msrc/home/admin/scripts/keytab_admin.sh Keytab entry for host "t01n01.emsl.pnl.gov"? (y, n, q) [y] - domain name 'nwtest' is left off. Be sure to correct the machine name if needed.

(3) Install secure shell **/afs/msrc/wsadmin/share/bin/InstallSsh tail /afs/msrc/wsadmin/share/log/'hostname'.log** (to check the success of the configuration)
(4) Add auto-startup for sshd to /etc/rc.local

78. *Configure cfengine*

79. *Add users*

a) cp -p /systems/config/lib/security/mkuser.default /lib/security/mkuser.default
b) cp -p /systems/config/lib/security/mkuser.sys /lib/security/mkuser.sys
c) cp -p /systems/cfg/etc/passwd.cw1 /etc
d) cp -p /systems/cfg/etc/security/passwd.cw1 /etc/security/passwd
e) cp -p /systems/cfg/etc/group.cw1 /etc/group
f) cp -p /systems/cfg/etc/security/user.cw1 /etc/security/user
g) cp -p /systems/cfg/etc/security/login.cfg.cw1 /etc/security/login.cfg
h) vi /etc/security/user
i) Make sure the default stanza is Change the default stanza to *SYSTEM = "compat"*

VI. References

1. *http://www.rs6000.ibm.com/resource/aix_resource/sp_books/index.html*
2. http://www.rs6000.ibm.com/resource/hardware_docs/ *(For Hardware related questions)*
3. http://www.redbooks.ibm.com
4. Parallel System Support Programs (PSSP) for AIX, Version 3 Release 1 Documentation (November 1998), Installation and Migration Guide (GA22-7347), Administration Guide (SA22-7348), Managing Shared Disks (SA22-7349)
5. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.1 Step 1: Update the root User Path
6. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.2 Verify the Control Workstation Requirements
7. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.5 Step 5: Configure RS-232 Control Lines
8. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.6 Step 6: Tune All Control Workstation Network Adapters
9. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.7 Step 7: Configure the Control Workstation Ethernet Adapters
10. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.9 Step 9: Ensure that the Necessary Daemons Are Running on the Control Workstation
11. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.10 Step 10: Change the Control Workstation Maximum Default Processes
12. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.13 Step 13: Define Space for the /spdata Directory
13. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.14 Step 14: Create the Required /spdata Directories
14. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.2.15 Step 15: Copy the AIX LPP Images and Other Required AIX LPPs and PTFs
15. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.3 Task B. Install PSSP on the Control Workstation 2.3.1 Step 16: Copy the PSSP Images 2.3.2 Step 17: Copy a Basic AIX (mksysb) Image 2.3.3 Step 18: Install PSSP on the Control Workstation 2.3.4 File Sets Installed on the Control Workstation 2.3.5 Installation without AIX Preinstalled 2.3.6 PSSP Installation Instructions 2.3.7 Step 19: Initialize RS/6000 SP Authentication Services 2.3.8 Step 20: Complete System Support Installation on the Control Workstation 2.3.9 Step 21: Add the PSSP T/EC Adapter 2.3.10 Step 22: Run SDR and System Monitor Verification Tests
16. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.3.1.1 Step 16.1: Copy PSSP Images from Media
17. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.3.7 Step 19: Initialize RS/6000 SP Authentication Services
18. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.3.8 Step 20: Complete System Support Installation on the Control Workstation
19. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.3.10 Step 22: Run SDR and System Monitor Verification Tests
20. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.1 Step 23: Enter Site Environment Information

21. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.2 Step 24: Enter Frame Information and Reinitialize the SDR
22. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.3 Step 25: Verify System Monitor Installation
23. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.4 Step 26: Verify Frame Information
24. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.5 Step 27: Update the State of the Supervisor Microcode
25. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.6 Step 28: Enter the Required Node Information
26. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.7 Step 29: Acquire the Hardware Ethernet Addresses
27. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.8 Step 30: Verify that the Ethernet Addresses Were Acquired
28. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.9 Step 31: Configure Additional Adapters for Nodes
29. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.10 Step 32: Configure Initial Host Names for Nodes
30. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.11 Step 33: Create Authorization Files
31. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.12 Step 34: Enable Selected Authentication Methods
32. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.14 Step 36: Start System Partition-Sensitive Subsystems
33. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.15 Step 37: Verify that System Partition-Sensitive Subsystems Have Started
34. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.17 Step 39: Set Up Nodes to Be Installed
35. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.18 Step 40: Verify All Node Information
36. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.21 Step 43: Configure the Control Workstation as the Boot/Install Server
37. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.5.2 Step 45: Verify that the System Management Tools Were Correctly Installed
38. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.4.21 Step 46: Set Up the Switch Subtopics: 2.5.3.1 Step 46.1: Select a Topology File 2.5.3.2 Step 46.2: Managing the Switch Topology Files 2.5.3.3 Step 46.3: Annotating a Switch Topology File 2.5.3.4 Step 46.4: Storing the Switch Topology File in the SDR
39. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.5.4 Step 47: Verify the Switch Primary and Primary Backup Nodes
40. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.5.5 Step 48: Set the Switch Clock Source for All Switches
41. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.6.1 Step 50: Network Boot Optional Boot/Install Servers
42. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.6.4 Step 53: Verify Node Installation
43. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.6.5 Step 54: Run Verification Tests on All Nodes

44. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.6.8 Step 57: Start the Optional Switch
45. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.6.9 Step 58: Verify that the Switch Was Installed Correctly
46. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/sspim310/2.6.10 Step 59: Tune the Network Adapters
47. General Parallel File System for AIX, Version 1 Release 2 Documentation (November 1998), Installation and Administration (SA22-7278-02)
48. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst09.html#HDRINSTEP1 Step 1. Installation Procedure
49. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst09.html#HDRINSTEP2 Step 2. Verify GPFS Installation
50. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst09.html#HDRINSTEP3 Step 3. Tune Switch
51. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst09.html#HDRKERB Step 4. Configure sysctl
52. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst09.html#HDRINSTEP5 Step 5. Tune IBM Virtual Shared Disk
53. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst12.html#HDRVSDTUN "IBM Virtual Shared Disk Parameters and GPFS Performance"
54. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst09.html#Header_75 Step 6. Reboot
55. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRADMIN Chapter 5. Performing GPFS Administration Tasks
    - "Managing File Systems"
    - "Managing Disks"
    - "Managing Nodes"
    - "Managing GPFS Quotas"
    - "Managing Access Control Lists"
56. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRCONSMIT Configure Using SMIT
57. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRCFGCMD Configure Using mmconfig Command
58. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRSTARSTO Starting and Stopping GPFS
59. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRFSMANAG Managing File Systems
    - "Creating a File System"
    - "Building Disk Descriptors"
    - "Deleting a File System"
    - "Modifying File System Attributes"
60. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst08.html#HDRSAMPVSD "Planning Sample IBM Virtual Shared Disks"
61. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRDELETFS Deleting a File System
62. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRCHKFS Checking and Repairing a File System

63. http://www.rs6000.ibm.com/resource/aix_resource/sp_books/gpfs/install_admin/install_admin_v1r2/gpfs1mst11.html#HDRDIMANAG Managing Disks
    - "Adding Disks to a File System"
    - "Deleting Disks from a File System"
    - "Displaying GPFS Disk States"
    - "Changing GPFS Disk States and Parameters"
    - "Replacing Disks in a GPFS File System"
64. Loadleveler Version 2 Release 1 Documentation (November, 1998), Using and Administering (SA22-7311-00), Diagnosis and Messages (GA22-7277-00), Installation Memo (GI10-0642-00)
65. http://www.s390.ibm.com:80/bookmgr-cgi/bookmgr.cmd/BOOKS/LLIV2R1/CCONTENTSInstalling LoadLeveler
    - Step 1: Decide on Directories for LoadLeveler
    - Step 2: Login as root
    - Step 3: Create the loadl Group Name
    - Step 4: Create the loadl User ID
    - Step 5: Receive the Installation Image
    - Step 6: Run the Installation Script llinit
    - Step 7: Make the Graphical User Interface Resource File Available
    - Step 8: Update the PATH Environment Variable
    - Step 9: Make the Man Pages Available
    - Step 10: Repeat the Appropriate Steps for Each Machine
    - Step 11: Configure Your LoadLeveler System
66. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLIV2R1/3.6 Step 6: Run the Installation Script llinit
67. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLIV2R1/3.8 Step 8: Update the PATH Environment Variable
68. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLIV2R1/3.9 Step 9: Make the Man Pages Available
69. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLAV2R1/3.1
    - Overview
    - Planning Considerations
    - Quick Set Up
    - Administering LoadLeveler
    - Configuring LoadLeveler
    - Keyword Summary
70. http://www.tivoli.com/support/storage_mgr/pubs/admanual.htm (***Main Page***)
71. http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLAV2R1/APPENDIX1.3
72. APPENDIX1.3.1 Sample Administration File
    - APPENDIX1.3.2 Sample Configuration File
73. RS/6000 SP Performance Tuning Red Book, December 1998 (SG24-5340-00)

VII. Appendix

80. *Configure Tivoli Storage Manager* - NWMPP1 only

http://www.tivoli.com/support/storage_mgr/pubs/admanual.htm

81. *The "gender" file. The* gender *file tells the* customize *and* customize.local *scripts how to behave. Below is a sample* gender *file:*

Files are current as of 04/25/2000

```
################################################################################
#
#                       MSCF GENDERS FILE
#
################################################################################

################################################################################
#                               MPP1
################################################################################

et0101 GPFS_SERVER INSTALL_SERVER IP_FORWARD # Comments allowed at end of line
et0102 COMPUTE_NODE FS_CLIENT
et0103 COMPUTE_NODE FS_CLIENT
et0104. . . . et0916 COMPUTE_NODE FS_CLIENT
et1001 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et1002. . . . et1016 COMPUTE_NODE FS_CLIENT
et1101 COMPUTE_NODE FS_CLIENT IP_FORWARD
et1102. . . . et1116 COMPUTE_NODE FS_CLIENT
et1201 COMPUTE_NODE FS_CLIENT IP_FORWARD
et1202. . . . et1216 COMPUTE_NODE FS_CLIENT
et1301 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et1302. . . .  COMPUTE_NODE FS_CLIENT
et1401 FS_CLIENT GPFS_SERVER IP_FORWARD
et1403 FS_CLIENT GPFS_SERVER
et1405 FS_CLIENT GPFS_SERVER
et1407 FS_CLIENT ADSM_SERVER
et1409 FS_CLIENT INTERACTIVE ATM_NODE
et1411 FS_CLIENT INTERACTIVE ATM_NODE
et1413 FS_CLIENT
et1415 FS_CLIENT SCHEDD_NODE EPRIMARY
et1501 FS_CLIENT GPFS_SERVER
et1503 FS_CLIENT GPFS_SERVER
et1505 FS_CLIENT GPFS_SERVER
et1507 FS_CLIENT
et1509 FS_CLIENT
et1511 FS_CLIENT
et1513 FS_CLIENT
et1515 FS_CLIENT SCHEDD_NODE EPRIMARY
et1601 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et1602. . . . et1616 COMPUTE_NODE FS_CLIENT
et1701 COMPUTE_NODE FS_CLIENT IP_FORWARD
et1702. . . . et1716 COMPUTE_NODE FS_CLIENT
et1801 COMPUTE_NODE FS_CLIENT IP_FORWARD
et1802. . . . et1816 COMPUTE_NODE FS_CLIENT
et1901 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et1902. . . . et1916 COMPUTE_NODE FS_CLIENT
et2001 COMPUTE_NODE FS_CLIENT IP_FORWARD
et2002. . . . et2016 COMPUTE_NODE FS_CLIENT
```

```
et2101 COMPUTE_NODE FS_CLIENT IP_FORWARD
et2102. . . . et2116 COMPUTE_NODE FS_CLIENT
et2201 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et2202. . . . et2216 COMPUTE_NODE FS_CLIENT
et2301 COMPUTE_NODE FS_CLIENT IP_FORWARD
et2302. . . . et2316 COMPUTE_NODE FS_CLIENT
et2401 COMPUTE_NODE FS_CLIENT IP_FORWARD
et2402. . . . et2416 COMPUTE_NODE FS_CLIENT
et2501 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et2502 FS_CLIENT GPFS_SERVER
et2503. . . . et2516 COMPUTE_NODE FS_CLIENT
et2601 COMPUTE_NODE FS_CLIENT IP_FORWARD
et2602. . . . et2616 COMPUTE_NODE FS_CLIENT
et2701 COMPUTE_NODE FS_CLIENT IP_FORWARD
et2702. . . . et2716 COMPUTE_NODE FS_CLIENT
et2801 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et2802. . . . et2816 COMPUTE_NODE FS_CLIENT
et2901 COMPUTE_NODE FS_CLIENT IP_FORWARD
et2902. . . . et2916 COMPUTE_NODE FS_CLIENT
et3001 COMPUTE_NODE FS_CLIENT IP_FORWARD
et3002. . . . et3016 COMPUTE_NODE FS_CLIENT
et3101 GPFS_SERVER INSTALL_SERVER IP_FORWARD
et3102. . . . et3116 COMPUTE_NODE FS_CLIENT
et3201 COMPUTE_NODE FS_CLIENT IP_FORWARD
et3202. . . . et3216 COMPUTE_NODE FS_CLIENT
et3301 COMPUTE_NODE FS_CLIENT IP_FORWARD
et3302. . . . et3316 COMPUTE_NODE FS_CLIENT
################################################################################
#                               TEST
################################################################################

t01n01 COMPUTE_NODE FS_CLIENT INTERACTIVE SCHEDD_NODE EPRIMARY
t01n03. . . . t01n08 COMPUTE_NODE FS_CLIENT
t01n09 COMPUTE_NODE FS_CLIENT INTERACTIVE

################################################################################
#                               ECS1
################################################################################

x01n01 COMPUTE_NODE FS_CLIENT INTERACTIVE ATM_NODE SMP_NODE EPRIMARY
x01n02. . . . x01n07 COMPUTE_NODE FS_CLIENT SMP_NODE
x01n08 COMPUTE_NODE FS_CLIENT ATM_NODE SCHEDD_NODE SMP_NODE LICENSE_SERVER
x01n09. . . . x01n16 COMPUTE_NODE FS_CLIENT SMP_NODE
x02n01 COMPUTE_NODE FS_CLIENT INTERACTIVE SMP_NODE
x02n05 COMPUTE_NODE FS_CLIENT SMP_NODE
x02n09 COMPUTE_NODE FS_CLIENT SMP_NODE
x02n13 COMPUTE_NODE FS_CLIENT SMP_NODE
```

82. *The customize and customize.local scripts.*

These scripts are foundational to the success of the installation. Below are samples:

Customize

```
#!/bin/ksh
#
# This script is intended to customize a single node or a set of nodes (e.g. an
# entire frame or machine dictated by setting WCOLL) after the vanilla install
# It is to be run from the control workstation. See Usage().
```

```
#
#Copyright Notice
#
#Maui High Performance Computing Center-authored documents are sponsored in
#part by the United States Department of Defense through Phillips Laboratory,
#Air Force Matiriel Command, United States Air Force, under cooperative
#agreement number F29601-93-2-0001. Accordingly, the United States Government
#retains a nonexclusive, royalty-free license to publish or reproduce these
#documents, or allow others to do so, for United States Government purposes.
#All documents available from this server may be protected by United States
#and international copyright laws.
#
#This document is copyrighted by the Maui High Performance Computing Center
#(MHPCC). Educational institutions are encouraged to reproduce and distribute
#these materials for educational use. Please retain this copyright notice and
#include this statement with any copies that you make. Also, the MHPCC requests
#that you send notification of their use to help@mail.mhpcc.edu.
#
#Commercial use of these materials is prohibited without prior written
#permission. Contact help@mail.mhpcc.edu for more information.
#
#Copyright 1994, 1995, 1996, 1997 Maui High Performance Computing Center. All
#rights reserved.


##############################################################################
# Function abort
##############################################################################

abort()
{
print $1
print "Customize aborted ..."
exit -1
}

##############################################################################
# Functions to determine CW of Origin
##############################################################################

cw1()
{ [[ $CW = "cw1" ]] }

cw3()
{ [[ $CW = "cw3" ]] }

cw4()
{ [[ $CW = "cw4" ]] }

##############################################################################
# MAIN Main main
##############################################################################

if [[ $1 = "" ]]; then echo "USAGE: $0 $WCOLL\n"; exit -1; fi

export WCOLL=$1
RSH="dsh -f 16"
CP="/bin/cp"
```

```
RCP="/usr/lpp/ssp/rcmd/bin/rcp"
CW=`hostname -s`
PCP="/usr/lpp/ssp/bin/pcp"

if [ "${CW%?}" != "cw" ]
then abort "This command must be run from the control workstation."
fi

print "Verifying access to the nodes."
test -f $WCOLL || abort "Please specify a valid WCOLL file."

cat $WCOLL | $PCP -w - /systems/cfg/etc/genders /etc/genders
$RSH date >/dev/null 2>&1 || abort "DSH failed. You may need to kinit."
($RSH "ls -l /etc/genders 2>&1" | grep "not") \
&& abort "Failed to copy /etc/genders to some nodes."
print "/etc/genders copied to all nodes."

print "\nMake sure switch is up on all nodes..."

cat $WCOLL|sed 's/n/s/'|/systems/bin/fping -f - || abort "The switch is not up on
all nodes."

print "\nImplementing Routes\n"
      cat $WCOLL| pcp -w - /systems/cfg/etc/router /etc/router
      cat $WCOLL| pcp -w - /systems/cfg/etc/routes /etc/routes
$RSH "/etc/router"

print "\nMounting /systems on the nodes."
$RSH "mkdir /systems 2>/dev/null"
$RSH "[ `lsfs /systems | grep /systems | wc -l` = 0 ] || mount
198.129.136.5:/systems /systems"

print "Running Node-Specific Customize Portion"
print "Please wait . . . . You can tail -f /tmp/cust.out on any node"
$RSH "/systems/bin/customize.local"

# Customization Complete
print "Customization Completed"
print "Do Not Forget To REBOOT These Nodes!"
```

83. *Customize.local*

```
#!/bin/ksh
#
#Copyright Notice
#
#Maui High Performance Computing Center-authored documents are sponsored in
#part by the United States Department of Defense through Phillips Laboratory,
#Air Force Matiriel Command, United States Air Force, under cooperative
#agreement number F29601-93-2-0001. Accordingly, the United States Government
#retains a nonexclusive, royalty-free license to publish or reproduce these
#documents, or allow others to do so, for United States Government purposes.
#All documents available from this server may be protected by United States
#and international copyright laws.
#
```

```
#This document is copyrighted by the Maui High Performance Computing Center
#(MHPCC). Educational institutions are encouraged to reproduce and distribute
#these materials for educational use. Please retain this copyright notice and
#include this statement with any copies that you make. Also, the MHPCC requests
#that you send notification of their use to help@mail.mhpcc.edu.
#
#Commercial use of these materials is prohibited without prior written
#permission. Contact help@mail.mhpcc.edu for more information.
#
#Copyright 1994, 1995, 1996, Maui High Performance Computing Center. All
#rights reserved.
#
#Updated 09/07/99 mv
#Updated 09/08/99 mv, bp, rb, rw

##############################################################################
# function iam,iamnot -- determines what type of node you are
##############################################################################

function iam {
  /usr/bin/grep `hostname -s` /etc/genders | /usr/bin/sed 's/#.*//' | /usr/bin/grep
$1 >/dev/null 2>&1
  }

function iamnot {
  ! iam
  }

##############################################################################
# function nwxxxx -- determines what system you are on
# function cws -- determines if you are on a control workstation
# function cwx -- determines what control workstation you are on
##############################################################################

nwmpp1() { [[ $CWS = "cw1" ]] }
nwtest() { [[ $CWS = "cw3" ]] }
nwecs1() { [[ $CWS = "cw4" ]] }
cws() { [[ ${HOST%%[0-9]+} = "cw" ]] }
cw1() { [[ $HOST = "cw1" ]] }
cw3() { [[ $HOST = "cw3" ]] }
cw4() { [[ $HOST = "cw4" ]] }

##############################################################################
# Function abort
##############################################################################

abort()
{
print $1
print "Customize aborted ..."
exit -1
}
```

```
################################################################################
# Function copy
################################################################################

function copy {
# Create an original copy before clobbering
test -e $2 && test ! -a $2.orig && cp $2 $2.orig
# Only do the copy if the file size is different
[[ `ls -l $1 2>/dev/null |awk '{print $5}'` != `ls -l $2 2>/dev/null | \
awk '{print $5}'` ]] && cp -p $1 $2
}
################################################################################
#
#            Wrapping all system calls to echo command executed
################################################################################
function runit {
  print "*** executing: $1 ***"
  eval $1
  print "*** done: $1 RC=$? ***"
}
################################################################################
#     INITIALIZATION
################################################################################
# Initialize variables

DEBUG=0
HOST=`hostname -s`
CWS=`/usr/lpp/ssp/bin/SDRGetObjects -x SP control_workstation|awk '{print $1}'`
NODENUM=`/usr/lpp/ssp/install/bin/node_number`
FRAME=`/usr/lpp/ssp/bin/SDRGetObjects -x Node node_number==$NODENUM frame_number`
RCP="/usr/lpp/ssp/rcmd/bin/rcp -p"

# Get appropriate code level info.
nwmpp1 && AIX=432
nwtest && AIX=432
nwecs1 && AIX=433
nwmpp1 && SSP=31
nwtest && SSP=31
nwecs1 && SSP=311

echo "Beginning Local Customization for $HOST"
echo "tail -f /tmp/cust.out to track progress"

exec 1>/tmp/cust.out    # Direct Output to file
exec 2>&1         # Stderr to same file

# Make sure systems is mounted (mounting performed in parent)
mount | grep /systems || abort "systems not mounted"

# Get Kerberos tickets for RCP
/usr/lpp/ssp/rcmd/bin/rcmdtgt # Getting kerberos tickets for rcp

# SPAM ALERT -- Don't run on CWS (yet)
cws && abort "Do not run $0 on the control workstation!"

# Copy sm_inst hacks needed to install ptfs and lpps
copy /systems/cfg/usr/lib/instl/sm_inst /usr/lib/instl/sm_inst
```

```
# Ensure genders file exists and is nonzero size
[[ ! -s /etc/genders ]] && abort "NO /etc/genders FILE FOUND!!!"

################################################################################
# MAIN -- Beginning of Node Customization
################################################################################

echo "\nCUSTOMIZING NODE: $HOST [$CWS]"

################################################################################
#     DUMP SPACE
################################################################################
#
#
echo "\nCreating new Dump Space."
DUMPSZ=`sysdumpdev -e | awk '{print $7}'`
# Get Physical Partition size in MB
let "PPSZ=`lsvg rootvg | grep 'PP SIZE' | awk '{print $6}'"
# Calculate the PP size in bytes.
let "PPSZBYTES=$PPSZ*1048576"
# Round up to next integer.
let "LVPPBYTES=$DUMPSZ*100/$PPSZBYTES+50"
# Convert back to number of PPs.
let "TOTPP=$LVPPBYTES/100"
# Double the current dump size to compensate for quiet system.
let "PPDUMPSZ=$TOTPP*2"
runit "mklv -y'dumplv' rootvg $PPDUMPSZ"
runit "sysdumpdev -P -p /dev/dumplv"

################################################################################
#       INSTALL AFS
################################################################################
if iam INTERACTIVE
then
# DO NOT RE-UNTAR AFS FILES IF AFS IS ALREADY ACTIVE UNLESS YOU LIKE 888s
if ( test ! -e /etc/rc.afs )
then
      mkdir /usr/vice
      mkdir /afs
      # Expand /usr as necessary
      usrfree=`df -v /usr | tail -1 | awk '{ print $4 }'`
      if [[ $usrfree -lt 40000 ]]
      then
            let "usradd=40000-$usrfree"
            chfs -a size=+$usradd /usr
      fi
      zcat /systems/src/emsl-aix43-afs-3.4_5.67.tar.Z|tar xvf -
      # Add a line that shouldn't be needed if AFS really was for AIX 432.
      # This sets the AFS sysname to access aix43 namespace
      echo "/usr/afsws/bin/fs sysname rs_aix43" >> /etc/rc.afs
      # Create AFS cache filesystem
      crfs -v jfs -m'/usr/vice/cache' -A'yes' -p'rw' -t'no' -a size=1179648 -g
rootvg
      # Install AFS cacheinfo file that sets cache size to 500MB
      cp -p /systems/cfg/usr/vice/etc/cacheinfo /usr/vice/etc/cacheinfo
      # AFS needs a vfs entry.
      echo "afs     4       none            none" >> /etc/vfs
      # Mount cachefs now to change perms
```

```
      mount /usr/vice/cache &&  chmod 700 /usr/vice/cache && chmod g-s
/usr/vice/cache
      # Add entry into inittab to autostart AFS at next reboot
      mkitab "rcafs:2:wait:/etc/rc.afs > /dev/console 2>&1 # Start AFS cache
manager"
fi
fi
#############################################################################
#     LVM -- PHYSICAL VOLUMES/STRIPING/VOLUME GROUPS
#############################################################################

# extend rootvg by including non-ssa hdisk1 if it exists
if iam FS_CLIENT
then
lspv |grep hdisk1|grep None && lsdev -Cc disk|grep hdisk1 |grep -v -i -e ssa -e
other > /dev/null && runit "extendvg -f rootvg hdisk1";lspv
lspv |grep hdisk2|grep None && lsdev -Cc disk|grep hdisk2 |grep -v -i -e ssa -e
other > /dev/null && runit "extendvg -f rootvg hdisk2";lspv
lspv |grep hdisk3|grep None && lsdev -Cc disk|grep hdisk3 |grep -v -i -e ssa -e
other > /dev/null && runit "extendvg -f rootvg hdisk3";lspv
fi
#############################################################################
#     PAGING SPACE
#############################################################################
# mkps command  mkps -s'2' -n'' -a'' rootvg  the -n'' make active now
echo "\nCreating Additional Paging Space."
PPSIZE=`lsvg rootvg | grep "PP SIZE"|awk '{print $6}'`
PAGESP=`lsps -a|awk '{print $4}'|tail +2|tr 'MB' '+.'|bc|tail -1`
let "MEMORY=`bootinfo -r`/1024"
      case $MEMORY in
      128)  if [[ $PAGESP -lt 256 ]]
            then runit "mkps -s $(( (256-$PAGESP)/$PPSIZE )) -a rootvg"
            fi;;
      256)  if [[ $PAGESP -lt 512 ]]
            then runit "mkps -s $(( (512-$PAGESP)/$PPSIZE )) -a rootvg"
            fi;;
      512)  if [[ $PAGESP -lt 512 ]]
            then runit "mkps -s $(( (840-$PAGESP)/$PPSIZE )) -a rootvg"
            fi;;
      1024) if [[ $PAGESP -lt 1024 ]]
            then runit "mkps -s $(( (1024-$PAGESP)/$PPSIZE )) -a rootvg"
            fi;;
      *)    if [[ $PAGESP -lt $(($MEMORY)) ]]
            then runit "mkps -s $(( ($MEMORY-$PAGESP)/$PPSIZE )) -a rootvg"
            fi;;
      esac
runit "lsps -a"
#############################################################################
#     SCRATCH
#############################################################################

if ( iam FS_CLIENT )
then
      echo "\nCreating /scratch."
      DISK=`bootinfo -s hdisk0`
        PPSIZE=`lsvg rootvg | grep "PP SIZE"|awk '{print $6}'`
      if (! lsfs /scratch 1>/dev/null 2>&1)
      then
```

```
            case $DISK in
            1920) runit "mklv -y scratchlv -S64K rootvg $(( 2048/$PPSIZE )) hdisk0
hdisk1";;
            4303) runit "mklv -y scratchlv -S64K rootvg $(( 4096/$PPSIZE )) hdisk0
hdisk1";;
            8678) runit "mklv -y scratchlv -S64K rootvg $(( 8192/$PPSIZE )) hdisk0
hdisk1";;
            17357)      runit "mklv -y scratchlv -S64K rootvg $(( 16384/$PPSIZE ))
hdisk0 hdisk1";;
            *)    runit "mklv -y scratchlv -S64K rootvg $(( $DISK/$PPSIZE )) hdisk0
hdisk1";;
            esac
            runit "crfs -v jfs -a bf=true -dscratchlv -m/scratch -Ay -prw"
      fi
      sleep 1
      runit "mount /scratch; df /scratch"
      chmod 777 /scratch
fi

################################################################################
#     INSTALL BASIC FILESETS
################################################################################

print "\nInstalling Basic filesets..."
runit "/systems/lpp/aix$AIX/install.aix"
nwmpp1 && runit "/systems/ptf/aix43/install.update"
nwtest && runit "/systems/ptf/aix43/install.update"
nwecs1 runit "/systems/ptf/aix433/install.update"
runit "/systems/lpp/ssp$SSP/install.ssp"
runit "/systems/lpp/ppe24/install.ppe"
runit "/systems/lpp/xlhpf14/install.rte"
runit "/systems/lpp/xlf610/install.rte"
runit "/systems/lpp/pessl31/install.pessl"
runit "/systems/lpp/gpfs/install.gpfs"

################################################################################
#       LOGIN NODES (install X11, etc.)
################################################################################

if iam INTERACTIVE
then
      runit "/systems/lpp/aix$AIX/install.login"
#     runit "/systems/lpp/graphics/install.graphics"
      runit "/systems/lpp/ssp$SSP/install.login"
      runit "/systems/lpp/ppe24/install.login"
      runit "/systems/lpp/vacpp40/install.cpp"
      runit "/systems/lpp/vac44/install.vac"
      runit "/systems/lpp/xlhpf14/install.cmp"
      runit "/systems/lpp/xlf610/install.xlf"
      runit "/systems/lpp/cxx364/install.cxx364"
      runit "/systems/lpp/vast/install.vast"
      runit "/systems/lpp/doc$AIX/install.man"
      runit "/systems/lpp/netscape/install.netscape"
      runit "/systems/lpp/perf/install.perf"
fi

################################################################################
#     GENERAL CUSTOMIZATION
```

```
################################################################################

# Modify inittab to call rc.local and to not start infod
mkitab "rclocal:2:once:/etc/rc.local"
# Modify inittab to start routes
mkitab "redoroutes:2:once:/usr/local/sbin/redo.routes > /dev/null"
## Make sure we have the right rc.local
copy /systems/cfg/etc/rc.local /etc/rc.local
nwmpp1 && copy /systems/cfg/etc/environment /etc/environment
nwtest && copy /systems/cfg/etc/environment /etc/environment
nwecs1 && copy /systems/cfg/etc/environment.ecs1 /etc/environment
copy /systems/cfg/etc/csh.cshrc.mpp1 /etc/csh.cshrc
rmdev -dl gxme0   2>/dev/null # Prevent cfgmgr errors
rmdev -dl rcm0    2>/dev/null # Prevent cfgmgr errors
rmdev -dl lft0    2>/dev/null # Prevent cfgmgr errors
/usr/lib/instl/sm_inst installp_cmd -u -f'devices.graphics.com' '-g'

################################################################################
#       COMPILERS/LICENSES/LIBRARIES
################################################################################

#Change AIX Licensing
runit "chlicense -u 65"

################################################################################
#       SYSTEM TUNING
################################################################################

## Where need aio?
runit "chdev -l aio0 -P -a autoconfig=available"
## Change maximum number of processes
runit "chdev -l sys0 -a maxuproc=512"


################################################################################
#     NETWORK CUSTOMIZATION
################################################################################

echo "\nNETWORK CUSTOMIZATION."
for i in `lscfg | grep " ent" | sed -e 's/^.*ent/ent/' -e 's/ .*//'`
do
    nwmpp1 && runit "chdev -l $i -a xmt_que_size=150 -P"
    nwtest && runit "chdev -l $i -a xmt_que_size=512 -P"
    nwecs1 && runit "chdev -l $i -a tx_que_size=256 -P"
done
# Change switch tuning to 16 MB's based on GPFS requirements
runit "/usr/lpp/ssp/css/chgcss -l css0 -a rpoolsize=16777216 -a spoolsize=16777216"
copy /systems/cfg/etc/netsvc.conf /etc/netsvc.conf
nwmpp1 && copy /systems/cfg/etc/resolv.conf.mpp1 /etc/resolv.conf
nwtest && copy /systems/cfg/etc/resolv.conf.nwtest /etc/resolv.conf
nwecs1 && copy /systems/cfg/etc/resolv.conf.ecs1 /etc/resolv.conf
nwecs1 && copy /systems/cfg/etc/ntp.conf.server.ecs1 /etc/ntp.conf
nwecs1 && copy /systems/cfg/etc/ntp.conf.client.ecs1 /etc/ntp.conf
nwmpp1 && copy /systems/cfg/etc/ntp.conf.server.mpp1 /etc/ntp.conf
nwmpp1 && copy /systems/cfg/etc/ntp.conf.client.mpp1 /etc/ntp.conf
nwtest && copy /systems/cfg/etc/ntp.conf.client.nwtest /etc/ntp.conf
copy /systems/cfg/etc/router /etc/router
copy /systems/cfg/etc/routes /etc/routes
```

```
copy /systems/cfg/etc/mounter /etc/mounter
copy /systems/cfg/etc/dce_er /etc/dce_er
copy /systems/cfg/etc/loadler /etc/loadler

################################################################################
#     SECURITY CUSTOMIZATION
################################################################################

#print "\nCopying over Security Files..."
copy /systems/cfg/etc/hosts /etc/hosts
copy /systems/cfg/etc/security/limits /etc/security/limits
copy /systems/cfg/etc/security/login.cfg /etc/security/login.cfg
chmod 640 /etc/security/limits /etc/security/login.cfg
copy /systems/cfg/profile.node /.profile
## Has to handle .rhosts hosts.equiv ...
copy /systems/cfg/.k5login /.k5login
chmod 600 /.k5login
copy /systems/src/tcp_wrappers_7.6/tcpd /usr/sbin/tcpd
runit "chmod 755 /usr/sbin/tcpd"
runit "chown root.system /usr/sbin/tcpd"
copy /systems/cfg/etc/hosts.allow /etc/hosts.allow
runit "mkdir -p /etc/banners"
copy /systems/cfg/etc/banners/* /etc/banners
copy /systems/cfg/etc/motd /etc/motd
copy /systems/cfg/etc/sshd_config /etc/sshd_config
runit "refresh -s inetd"


################################################################################
#       FILESYSTEMS
################################################################################

runit "/usr/sbin/chfs -a size=131072 /var"
runit "/usr/sbin/chfs -a size=131072 /"
runit "crfs -v jfs -g rootvg -a size=131072 -m /usr/local -A y -p rw -t n"
runit "mount /usr/local"      # Possibly wrap with test
#Mount /systems from prophet.emsl.pnl.gov
runit "/usr/sbin/mknfsmnt -f /systems -d /systems -h prophet -I -y -z"
#Mount home directory /u1 from neo.emsl.pnl.gov
runit "/usr/sbin/mknfsmnt -f /u1 -d /u1 -h neo -m nfs -a -t rw -w bg -S"
#Mount home directory /u2 from morpheus.emsl.pnl.gov
runit "/usr/sbin/mknfsmnt -f /u2 -d /u2 -h morpheus -m nfs -a -t rw -w bg -S"
runit "mount /u1"
runit "mount /u2"


################################################################################
#       INSTALL CFENGINE
################################################################################

test -e /usr/local/sbin || mkdir -p /usr/local/sbin
test -e /usr/local/sbin/cfengine || copy /systems/cfg/usr/local/sbin/cfengine
/usr/local/sbin/cfengine
test -e /usr/local/sbin/cfd || copy /systems/cfg/usr/local/sbin/cfd
/usr/local/sbin/cfd
test -e /usr/local/sbin/cfrun || copy /systems/cfg/usr/local/sbin/cfrun
/usr/local/sbin/cfrun
nwecs1 && copy /systems/cfg/etc/cfengine.conf.ecs1 /etc/cfengine.conf
```

```
nwtest && copy /systems/cfg/etc/cfengine.conf.test /etc/cfengine.conf
nwmpp1 && copy /systems/cfg/etc/cfengine.conf.mpp1 /etc/cfengine.conf
nwecs1 && copy /systems/cfg/etc/cfd.conf.ecs1 /etc/cfd.conf
nwtest && copy /systems/cfg/etc/cfd.conf.test /etc/cfd.conf
nwmpp1 && copy /systems/cfg/etc/cfd.conf.mpp1 /etc/cfd.conf
nwecs1 && copy /systems/cfg/etc/cfrun.hosts.ecs1 /etc/cfrun.hosts
nwtest && copy /systems/cfg/etc/cfrun.hosts.test /etc/cfrun.hosts
nwmpp1 && copy /systems/cfg/etc/cfrun.hosts.mpp1 /etc/cfrun.hosts
test -e /usr/local/sbin/cfengine && runit "export CFINPUTS=/etc;
/usr/local/sbin/cfengine -Dfull"

################################################################################
#     STATISTICS AND MONITORING
################################################################################

if iam INTERACTIVE
then
      copy /systems/cfg/etc/syslog.conf.login /etc/syslog.conf
else
      copy /systems/cfg/etc/syslog.conf.compute /etc/syslog.conf
fi

runit "refresh -s syslogd"

################################################################################
#     AUDITING
################################################################################

################################################################################
#     INETD
################################################################################

echo "\nCustomizing /etc/inetd.conf"
if iam INSTALL_SERVER
then
copy /systems/cfg/etc/inetd.conf.install_server /etc/inetd.conf
else
    copy /systems/cfg/etc/inetd.conf /etc/inetd.conf
fi

################################################################################
#     SSA
################################################################################

################################################################################
#     GPFS
################################################################################

$RCP $CWS:/etc/sysctl.acl /etc/sysctl.acl
$RCP $CWS:/etc/sysctl.vsd.acl /etc/sysctl.vsd.acl
$RCP $CWS:/etc/sysctl.pman.acl /etc/sysctl.pman.acl
$RCP $CWS:/etc/sysctl.mmcmd.acl /etc/sysctl.mmcmd.acl
$RCP $CWS:/etc/sysctl.rootcmds.acl /etc/sysctl.rootcmds.acl
$RCP $CWS:/etc/sysctl.conf /etc/sysctl.conf

################################################################################
#     STORAGE -- EMASS / SDM
################################################################################
```

```
################################################################################
#     NFS
################################################################################
# Will NFS Servers be external to the system?

if iam NFS_SERVER
  then
    echo "Configuring to be NFS SERVER.\n"
    runit "/usr/etc/chnfs -n 256 -b 16 -B"
    runit "/bin/chssys -s nfsd -a 256"
  else
    runit "/usr/etc/chnfs -b 24 -n 1 -B"
fi

################################################################################
#     LOADLEVELLER
################################################################################

if ( iam COMPUTE_NODE )
  then
    echo "\nCreating \/loadl filesystem"
    # Create /loadl filesystem
    runit "crfs -vjfs -g rootvg -m /loadl -Ay -prw -a size=98304"
    runit "mount /loadl"
fi

if ( ( iam SCHEDD_NODE || iam INTERACTIVE ) )
  then
    echo "\nCreating \/loadl filesystem"
    # Create /loadl filesystem
    runit "crfs -vjfs -g rootvg -m /loadl -Ay -prw -a size=524288"
    runit "mount /loadl"
fi

if ( ( iam COMPUTE_NODE || iam SCHEDD_NODE || iam INTERACTIVE ) )
  then
    echo "\nInstalling and Configuring LoadLeveller"

# Install LoadLeveller
    runit "/systems/lpp/loadl21/install.loadl"

# Create loadl group and loadl user
# The commands below need to be run on the CWS then supper needs to be run
#
lsgroup loadl 2>/dev/null || mkgroup -'a' id='11' loadl
lsuser -a loadl 2>/dev/null || mkuser -a id='204' pgrp='loadl' home='/loadl' loadl
chown loadl:loadl /loadl

runit "chown loadl.loadl /loadl"
nwtest && runit "su - loadl -c /usr/lpp/LoadL/full/bin/llinit -local /loadl -
release /usr/lpp/LoadL/full -cm cw3.nwtest.emsl.pnl.gov"
nwecs1 && runit "su - loadl -c /usr/lpp/LoadL/full/bin/llinit -local /loadl -
release /usr/lpp/LoadL/full -cm cw4.ecs1.emsl.pnl.gov"
nwmpp1 && runit "su - loadl -c /usr/lpp/LoadL/full/bin/llinit -local /loadl -
release /usr/lpp/LoadL/full -cm cw1.nwmpp1.emsl.pnl.gov"
runit "errupdate -f /usr/lpp/LoadL/full/include/loadl_err.S"
# Copy in Admin and LocalConfig
```

```
nwecs1 && copy /systems/cfg/loadl/LoadL_config.ecs1 /loadl/LoadL_config
nwecs1 && copy /systems/cfg/loadl/LoadL_config.local.ecs1 /loadl/LoadL_config.local
nwecs1 && copy /systems/cfg/loadl/LoadL_admin.ecs1 /loadl/LoadL_admin
nwmpp1 && copy /systems/cfg/loadl/LoadL_config.mpp1 /loadl/LoadL_config
nwmpp1 && copy /systems/cfg/loadl/LoadL_config.local.mpp1 /loadl/LoadL_config.local
nwmpp1 && copy /systems/cfg/loadl/LoadL_admin.mpp1 /loadl/LoadL_admin

# Copy in prolog, epilog, etc.
copy /systems/cfg/loadl/bin/prolog /loadl/bin/prolog
copy /systems/cfg/loadl/bin/epilog /loadl/bin/epilog
copy /systems/cfg/loadl/bin/prolog.x /loadl/bin/prolog.x
copy /systems/cfg/loadl/bin/epilog.x /loadl/bin/epilog.x
copy /systems/cfg/loadl/bin/prolog.sh.x.nwtest /loadl/bin/prolog.sh.x
copy /systems/cfg/loadl/bin/epilog.sh.x.nwtest /loadl/bin/epilog.sh.x

# Enable Log Saving Mechanism
copy /systems/cfg/loadl/bin/savelogs /loadl/bin/savelogs
grep savelogs /var/spool/cron/crontabs/root || (crontab -l;echo "5,15,25,35,45,55 *
* * * /loadl/bin/savelogs >/dev/null 2>/dev/null") | crontab

# Force Interactive Jobs to use LoadLeveller as Resource Manager
nwecs1 && copy /systems/cfg/etc/poe.limits /etc/poe.limits
nwtest && copy /systems/cfg/etc/poe.limits /etc/poe.limits
nwmpp1 && copy /systems/cfg/etc/poe.limits /etc/poe.limits

# Mauiisms
mkdir -p /loadl/maui/bin 2>/dev/null
#copy /systems/cfg/loadl/maui/bin/wakeup /loadl/maui/bin/wakeup
#copy /systems/cfg/loadl/maui/bin/tellme /loadl/maui/bin/tellme
#copy /systems/cfg/loadl/maui/maui.cfg /loadl/maui/maui.cfg
#copy /systems/cfg/loadl/maui/bin/schedctl /loadl/maui/bin/schedctl
$RCP -p $CWS:/systems/cfg/loadl/maui/bin/wakeup /loadl/maui/bin/wakup
$RCP -p $CWS:/systems/cfg/loadl/maui/bin/tellme /loadl/maui/bin/tellme
$RCP -p $CWS:/systems/cfg/loadl/maui/maui.cfg /loadl/maui/maui.cfg
$RCP -p $CWS:/systems/cfg/loadl/maui/bin/schedctl /loadl/maui/bin/schedctl
#$RCP -p $CWS:/loadl/maui/maui.cfg /loadl/maui/maui.cfg
#$RCP -p $CWS:/loadl/maui/bin/sched /loadl/maui/bin/sched

fi

if iam SCHEDD_NODE
then
      nwtest && copy /systems/cfg/loadl/LoadL_config.local.sched.nwtest
/loadl/LoadL_config.local
      nwecs1 && copy /systems/cfg/loadl/LoadL_config.local.sched.ecs1
/loadl/LoadL_config.local
      nwmpp1 && copy /systems/cfg/loadl/LoadL_config.local.sched.mpp1
/loadl/LoadL_config.local
fi

##############################################################################
#     HIPPI
##############################################################################

##############################################################################
#     SMP
##############################################################################
```

```
#if iam SMP_HOST
#  then
#echo "Customizing for SMP_HOST"
#/systems/lpp/xlfbeta/install.script
#cp /systems/cfg/etc/xlC.cfg.SMP /etc/xlC.cfg
#fi

################################################################################
#     SYMBOLIC LINKS
################################################################################

################################################################################
#     APPLICATIONS -- LOCAL AND REMOTE
################################################################################

################################################################################
#     SENDMAIL
################################################################################

#print "\nCustomizing sendmail"
# Talk to Bill
#"test -e /etc/sendmail.cf.sav || cp /etc/sendmail.cf /etc/sendmail.cf.sav"
#"cp /systems/cfg/etc/sendmail.cf /etc/sendmail.cf"
#"chmod o-r /etc/sendmail.cf"

################################################################################
#     WHACKER
################################################################################

################################################################################
#       GATEWAY NODES (need ATM filesets, etc.)
################################################################################

################################################################################
#       ADSM
################################################################################

################################################################################
#       WEB SERVICES
################################################################################

################################################################################
#       LICENSE SERVICE
################################################################################

#/usr/ibmcxx/bin/replaceCSET  # (run twice - once for exec links once for man
pages)
#/usr/ibmcxx/bin/replaceCSET
#        o rcp /var/ifor/i4ls.ini cw4:/systems/cfg//var/ifor/i4ls.ini.ecs
runit "mkdir /usr/lib/vastls/"
nwecs1 && grep -p'# -' `hostname` /systems/cfg/usr/lib/vastls/vast2x.cfg.ecs1 >
/usr/lib/vastls/vast2x.cfg
nwtest && grep -p'# -' `hostname` /systems/cfg/usr/lib/vastls/vast2x.cfg.nwtest >
/usr/lib/vastls/vast2x.cfg
runit "chmod 755 /usr/lib/vastls/vast2x.cfg"
```

```
################################################################################
#       INSTALL DCE
################################################################################

runit "/systems/lpp/dce22/install.dce"
runit "/systems/lpp/dce22/install.dce.priv"

################################################################################
#       AIX 432/433 FIXES - November 1999 Maintenance Release
################################################################################

print "\nApplying AIX updates..."
nwmpp1 && runit "/systems/ptf/aix43/install.update"
nwtest && runit "/systems/ptf/aix43/install.update"
nwecs1 && runit "/systems/ptf/aix433/install.update"
print "AIX update completed."

################################################################################
#     CONFIGURE DCE AND KERBEROS
################################################################################

#runit "crfs -vjfs -g rootvg -m /var/adm/dfs/cache -Ay -prw -a size=140000"
#runit "mount /var/adm/dfs/cache"
#runit "/usr/bin/cm sysname rs_aix43"           # dfs
test -e /usr/lpp/ssp/expect/bin || runit "mkdir -p /usr/lpp/ssp/expect/bin"
test -e /usr/lpp/ssp/expect/bin/expect || runit "$RCP -p
$CWS:/usr/lpp/ssp/expect/bin/expect /usr/lpp/ssp/expect/bin/expect"
test -e /usr/lpp/ssp/lib/tcl || runit "mkdir -p /usr/lpp/ssp/lib/tcl"
test -e /usr/lpp/ssp/lib/tcl/init.tcl || runit "$RCP -p
$CWS:/usr/lpp/ssp/lib/tcl/init.tcl /usr/lpp/ssp/lib/tcl/init.tcl"
runit "$RCP -p $CWS:/systems/bin/rmdce22.expect /tmp/rmdce22.expect"
runit "$RCP -p $CWS:/systems/bin/mkdce22.expect /tmp/mkdce22.expect"
runit "/tmp/rmdce22.expect"
runit "/tmp/mkdce22.expect"
runit "/usr/bin/rm -f /tmp/rmdce22.expect"
runit "/usr/bin/rm -f /tmp/mkdce22.expect"
copy /systems/cfg/dceunixd.sh /dceunixd.sh
################################################################################
#       INSTALL MASS LIBRARIES - Current Version is 2.5
################################################################################

# Installing as Root
#
# 1.login as root
# -- or --
# su to root
runit "cd /usr/lpp;tar -xvf /systems/lpp/mass/MASS_2_5_tar"
echo "MASS files will be restored to the directory /usr/lpp/mass."
runit "ln -s /usr/lpp/mass/libmass.a /usr/lib/libmass.a"
runit "ln -s /usr/lpp/mass/libmassv.a /usr/lib/libmassv.a"
runit "ln -s /usr/lpp/mass/libmassvp2.a /usr/lib/libmassvp2.a"
# (This step creates a symbolic link from the /usr/lpp/mass/libmass.a
# to /usr/lib/libmass.a, etc., so that users can specify the -lmass
# flag as a shorthand to link MASS routines.)
```

```
################################################################################
#       PSSP FIXES
################################################################################

print "\nApplying pssp fixes..."
runit "/systems/ptf/ssp31/install.update"
nwecs1 && runit "/systems/ptf/ssp311/install.update"
print "PSSP Upgrade Completed.\n"
if iam INTERACTIVE
then
runit "/usr/sbin/catman -w"
fi
copy /systems/cfg/usr/lib/security/mkuser.default /usr/lib/security/mkuser.default
/bin/lslpp -l | grep -i broken  >> /tmp/cust.out
```

84. Installing AFS

```
#!/bin/ksh
#
# AFS Installation on AIX 4.3.2
#
# by Bill Pitre and Scott Jackson
# last revised Feb 10, 1999
#
# Create /usr/vice if not already present.
ls -ld /usr/vice > /dev/null 2>&1 || mkdir /usr/vice
#
# Create /afs if not already present.
ls -ld /afs > /dev/null 2>&1 || mkdir /afs
#
# Provide enough space in which to install /usr/afsws
#
usrneeded=10600
usrfree=‘df -vk /usr | tail -1 | awk ’{ print $4 }’‘
if [[ $usrfree -lt $usrneeded ]]
then
chfs -a size=+21200 /usr
fi
zcat /systems/src/emsl-aix43-afs-3.4_5.67.tar.Z | tar xvf -
#
#    Create AFS cache logical volume and file system.
#    And then mount the file system.
#
#####mklv -y’afslv’ rootvg 8
#####crfs -v jfs -d’afslv’ -m’/usr/vice/cache’ -A’’‘locale yesstr | awk -F: ’{print $1}’‘’’ -p’rw’ -
t’’‘locale nostr | awk -F: ’{print $1}’‘’’ -a frag=’4096’ -a nbpi=’4096’ -a ag=’8’
# Make a 100MB AFS cache.
crfs -v jfs -m’/usr/vice/cache’ -A’’yes’’ -p’rw’ -t’’no’’ -a size=245760 -g rootvg
#
#####mount /dev/afslv /usr/vice/cache
mount /usr/vice/cache &&  chmod 700 /usr/vice/cache && chmod g-s /usr/vice/cache
```

```
mkitab "rcafs:2:wait:/etc/rc.afs > /dev/console 2>&1 # Start AFS cache manager"
cp -p /systems/cfg/usr/vice/etc/cacheinfo /usr/vice/etc/cacheinfo
nohup /etc/rc.afs &
```

85. *Unconfigure DCE Scripts*

```
#!/usr/lpp/ssp/expect/bin/expect
#uncfgdce22.expect
set password cum9tume
set timeout -1
set send_slow {1 .1}
eval spawn -noecho /systems/bin/uncfgdce.22
#eval spawn -noecho /tmp/catme
set done 0
while  {! $done } {
expect {
  "principal cell_admin:*" {
    send -s -- $password\r
    sleep .1
   }
  "*Enter to continue*" {
    send -s -- \r
    sleep .1
    set done 1
  }
   eof exit
  }
}
#
#     [Jan 19 1999, 15:59:59]
#
x() {
         /usr/bin/SmitPanels.tcl Smit_Commands_for_DCE_DFS unconfig -config_type full $*
         /usr/bin/chmod 700 /opt/dcelocal/tmp/dcedfscm
         /opt/dcelocal/tmp/dcedfscm
         /usr/bin/rm /opt/dcelocal/tmp/dcedfscm
         /usr/bin/dspmsg -s 1 dcesmitbase.cat 373 '          Press Enter to continue '
         /usr/bin/read string
        }
        x  'all' -cell_admin 'cell_admin' -pwdstr_principal 'pwd_strengthd'
```

86. Make DCE Scripts

```
#!/usr/lpp/ssp/expect/bin/expect
set password cum9tume
set timeout -1
set send_slow {1 .1}
eval spawn -noecho /systems/bin/mkdceclient.22
#eval spawn -noecho /tmp/catme
set done 0
```

```
while  {! $done } {
expect {
  "principal cell_admin:*" {
    send -s -- $password\r
    sleep .1
   }
  "*Enter to continue*" {
    send -s -- \r
    sleep .1
    set done 1
  }
   eof exit
  }
}
#!/bin/ksh
#
#     [Jan 06 1999, 18:07:44]
#
mkdir /var/adm/dfs
x(){
        /usr/bin/SmitPanels.tcl Smit_Commands_for_DCE_DFS config -config_type full $*
        /usr/bin/chmod 700 /opt/dcelocal/tmp/dcedfscm
        /opt/dcelocal/tmp/dcedfscm
        /usr/bin/rm /opt/dcelocal/tmp/dcedfscm
        /usr/bin/dspmsg -s 1 dcesmitbase.cat 373 '          Press Enter to continue '
        /usr/bin/read string
       }
       x  -cell_name '/.../dce.emsl.pnl.gov' -cell_admin 'cell_admin' -autostart 'yes' -protocol
'tcp_udp' -sec_master 'mohawk.emsl.pnl.gov' -cds_server 'skyhawk.emsl.pnl.gov' -lan_profile
'lan-profile' -cl_cache_size '70000' -cl_cache_dir '/var/adm/dfs/cache' rpc sec_cl cds_cl dfs_cl
ln -s /.../dce.emsl.pnl.gov/fs /dfs
```

87. *ntp configuration - May have been done through* customize
    a) cp /systems/cfg/etc/ntp.conf /etc/ntp.conf
    b) It may be necessary to kill an existing ntp process before the startsrc
    c) startsrc -s xntpd
    d) pcp -a /etc/ntp.conf /etc/ntp.conf
    e) dsh -af16 "startsrc -s xntpd"

88. *Configure nfs home everywhere else (compute+login nodes)* - done through customize
    a) dsh -a "/bin/rm -f /u;/bin/mkdir /u"
    b) smitty -xs /systems/tmp/nfs.mount mknfsmnt
        (1) * PATHNAME of mount point [/u1]
        (2) * PATHNAME of remote directory [/u1]
        (3) * HOST where remote directory resides [frame 15 node 7]
        (4) Mount type NAME []

(5) * Use SECURE mount option? No
(6) * MOUNT now, add entry to /etc/filesystems or both? Both
(7) * /etc/filesystems entry will mount the directory **no** on system RESTART.
(8) * MODE for this NFS file system read-write
(9) ATTEMPT mount in background or foreground background (Accept defaults from here down)

c) chmod +x /systems/tmp/nfs.mount
d) dsh -a "/systems/tmp/nfs.mount"
e) dsh -a "mount /u1; mount /u2;" # or appropriate directories.
f) Repeat above for /u2 on frame 15 node 13
g) **chmod +x /systems/tmp/nfs.mount**
h) dsh -a "/systems/tmp/nfs.mount"
i) dsh -a "mount /u1; mount /u2;" # or appropriate directories.

89. *Configure Loadleveler (FYI) - done on nodes through customize*
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLIV2R1/3.0

a) Configuration for the control workstation - Decide on directories for LoadLeveler. Use the information below for the location of the home and local directories. (***3.1 Step 1: Decide on Directories for LoadLeveler - LoadLeveler Installation MEMO***)
(1) Create /loadl filesystem at ~100MB's (NWTEST) on the cw; 434176 MB's (NWMPP1) *("customize" has already created this filesystem on the nodes check the size of /loadl on the nodes)*
(2) mount /loadl
(3) create loadl userid and group with uid 204 and gid 11 , home = /loadl
(a) mkgroup -'a' id='11' loadl
(b) mkuser -a id='204' pgrp='loadl' home='/loadl' loadl
(4) Add loadl as a kerberos principal -
(a) kadmin
(b) ank loadl (ank = add_new_key)
(5) telnet cw
(6) login as loadl and change passwd
(7) /usr/lpp/ssp/kerberos/bin/kinit loadl
(8) exit back to root
(9) run dinner to update the password file on the nodes
(10) Run the Installation Script ***llinit.***
http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLIV2R1/3.6
Step 6: Run the Installation Script llinit - LoadLeveler Installation MEMO) llinit does the following:
- Creates the LoadL_config file and the LoadL_admin file, if they don't already exist.
- Copies the LoadL_admin and the LoadL_config files from the release directory (in the samples subdirectory) into the home directory of loadl. Note: These files are a common resource for all the machines in the LoadLeveler cluster, and therefore must be made accessible to all members of the LoadLeveler pool.
- Sets local directory with permissions set to 775, 700, and 1777, respectively.

- Changing the Location of Directories: You can change the locations of these directories by changing the associated paths in the global configuration file. The global configuration file must reside in loadl's home directory or the location specified in /etc/LoadL.cfg. For example, if you want to move the log, spool, and execute directory from /loadl into /tmp/loadl, with appropriate permissions set, you can do so but you must create /tmp/loadl/spool, /tmp/loadl/execute, and /tmp/loadl/log in /tmp/loadl or LoadLeveler will not start up.
- Copies the LoadL_config.local file from the release directory (in the samples subdirectory) into the local directory.
- Creates symbolic links from the loadl home directory to the spool, execute, and log subdirectories and the LoadL_config.local file in the local directory (if home and local directories are not identical).

(11) You must perform this step using the LoadLeveler user ID. These instructions use loadl as this ID. To switch to the loadl ID, enter the following***:***

(12) su - loadl

(13) Ensure that your HOME environment variable is set to loadl's home directory***. - echo $HOME***

(14) cd /usr/lpp/LoadL/full/bin

(15) Check to make sure bin does not exist in /loadl. Enter the llinit command. For example, to run the llinit command with a local directory of /loadl, a release directory of /usr/lpp/LoadL/full, and a central manager named cw3.nwtest.emsl.pnl.gov, enter the following: **./llinit -local /loadl -release /usr/lpp/LoadL/full -cm cw3.nwtest.emsl.pnl.gov**

(a) Update the PATH Environment Variable. Add one of the following path statements to your PATH environment variable (http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLIV2R1/3.8 ***3.8 Step 8: Update the PATH Environment Variable - Loadleveler Installation MEMO***):

(b) /usr/lpp/LoadL/full/bin

(c) /loadl/bin

(16) Make the Man Pages Available

(a) /usr/lpp/LoadL/full/man/$LANG:/usr/lpp/LoadL/full/man

b) Configuration for the nodes (***customize*** *has already created a filesystem on the nodes, so make sure to check this out before continuing. If the filesystem is there, check ownership and contents before continuing,* skip steps 1-4 below and continue with step 5.) Check to make sure a file bin does not exist in /loadl. If it does, remove it.

(1) **smitty -xs create.loadlfs crjfsstd**

Volume group name rootvg
* SIZE of file system (in 512-byte blocks) [200000]
MOUNT POINT [/loadl]
Mount AUTOMATICALLY at system restart? Yes
chmod +x create.loadlfs

(2) **pcp -a create.loadlfs /create.loadlfs**

(3) **dsh -a "/create.loadlfs"**

(4) **dsh -a "mount /loadl"**
(5) **dsh -a "chown loadl.loadl /loadl"**
(6) **dsh -a "su - loadl -c 'cd /usr/lpp/LoadL/full/bin; ./llinit -local /loadl -release /usr/lpp/LoadL/full -cm cw3.nwtest.emsl.pnl.gov'"**
(7) Make sure there is a *prolog.x, prolog.sh.x, epilog.x, and epilog.sh.x in /loadl/bin/* and that they have the following permissions:
   (a) prolog.x, epilog.x *-rwsr-xr-t* (***chmod 05755 <file_name>***)
   (b) prolog.sh.x, epilog.sh. x -r-xr-x—x
(8) Make sure *ipforwarding=1* on frame1 node1.
(9) *4.4 Step 3. Update Your Configuration - Loadleveler Installation MEMO* http://ppdbooks.pok.ibm.com:80/cgi-bin/bookmgr/bookmgr.cmd/books/LLAV2R1/APPENDIX1.3This step must be performed as loadl. Ensure that your HOME environment variable is set to loadl's home directory.
   (a) Copy and rename the three files from the samples directory to the loadl home directory and the local directory. This step assumes the local directory is $HOME/hostname:
   cp /usr/lpp/LoadL/so/samples/LoadL_admin $HOME
   cp /usr/lpp/LoadL/so/samples/LoadL_config.so $HOME/LoadL_config
   cp usr/lpp/LoadL/so/samples/LoadL_config.local.so $HOME/hostname/LoadL_config.local
(10) Add /usr/lpp/LoadL/so/bin to the PATH environment variable.
(11) Add /usr/lpp/LoadL/so/man/$LANG:/usr/lpp/LoadL/so/man to the MANPATH environment variable.
(12) Create and update the *LoadL_admin* file by using the *llextSDR* command to create *machine stanzas* and *adapter stanzas* from the SDR (***llextSDR > /loadl/stanzas***). Make sure under *default: type = user* the default *class=batch* is in the *LoadL_admin* file.
(13) Update the *LoadL_config* file and ensure default *paths* are correct and that SCHEDULER_API = NO for testing purposes.
(14) Make sure the CWS and each node has the correct *LoadL_config.local* file. Below are samples:

```
schedd
START_DAEMONS   = TRUE
SCHEDD_RUNS_HERE  = TRUE
MAX_STARTERS    = 1
Class = {"batch"}
compute
START_DAEMONS   = TRUE
STARTD_RUNS_HERE = TRUE
SCHEDD_RUNS_HERE  = FALSE
MAX_STARTERS    = 1
Class = {"batch"}
```

(15) Start loadleveler and check its status
(16) llctl -g start
(17) llstatus
(18) Run a couple of test jobs