



SCO

UnixWare® 7 NonStop Clusters® Administration

Transparency Masters

Course Release: 1.0

For Software: UnixWare® 7 NonStop® Clusters 7.1.1

This publication is protected under copyright laws and international treaties.

Information in this document is subject to change without notice and does not represent a commitment on the part of The Santa Cruz Operation, Inc.

SCO, The Santa Cruz Operation, the SCO logos, VP/ix, SCO OpenServer, UnixWare, dbXtra, Panner, SCO Global Access, SCO MPX, SCO Doctor, SCO Doctor for Networks, SCO Doctor Lite, SCO Premier Motif, SCO Visual Tcl, SCO VisionFS, SCO ToolWare, SCO TermLite, SCO Vision 97, SCO CIFS Bridge, SCO WebTop, The Internet Way of Computing, and IWoC are trademarks or registered trademarks of The Santa Cruz Operation, Inc. in the USA and other countries. IXI, X.desktop, Deskworks, Wintif, IXI Panorama, Deskterm, and IXI Desktop are trademarks or registered trademarks of IXI Limited, a subsidiary of The Santa Cruz Operation, Inc. Visionware, Devkit.Vision, Esprit, Kodon, Super.Vision, Visionware Super.Vision, Term.Vision, Visionware Direction, the Visionware logo, Visionware SQL-Retriever, and XVision are registered trademarks of Visionware Limited, a subsidiary of The Santa Cruz Operation, Inc. SuperVision, TermVision, X-Visionware, SQL-Retriever, and Vision Builder are trademarks of Visionware Limited. UNIX is a registered trademark of The Open Group in the United States and other countries. Cheyenne and ARCserve are registered trademarks of Cheyenne Software, Inc. Netscape, Netscape Navigator, Netscape Communications Server, Netscape Commerce Server, Netscape Proxy Server, Netscape FastTrack Server, Netscape Enterprise Server, Netscape SuiteSpot, Netscape Catalog Server, Netscape News Server, Netscape Mail Server, and Netscape Navigator Gold are trademarks of Netscape Communications Corporation. NFS was developed by Legent Corporation (formerly Lachman Associates, Inc.) based on LACHMAN SYSTEM V NFS. LACHMAN is a trademark of Legent Corporation. NFS is a trademark of Sun Microsystems, Inc. TCP/IP was developed by Legent Corporation (formerly Lachman Associates, Inc.) based on LACHMAN SYSTEM V STREAMS TCP, a joint development of Lachman Associates and Convergent Technologies. MPX was developed by Corollary, Inc. VP/ix is a product developed and licensed by Phoenix Technologies, Ltd/INTERACTIVE Systems Corporation. XRemote is a registered trademark of Network Computing Devices, Inc. Oracle is a registered trademark of Oracle Corporation, Redwood City, California. Java is a trademark of Sun Microsystems, Inc. Compaq, Tandem, Integrity, NonStop, NetFlex, ProLiant, ROMPac, ServerNet, SmartStart, registered in United States Patent and Trademark Office. Intel and Pentium are registered trademarks, and Pentium II Xeon and Pentium III Xeon are trademarks of Intel Corporation. SCO, Veritas, VxVs, and VxVA are registered trademarks of Veritas Software Corporation. Other product names mentioned herein may be trademarks and/or registered trademarks of their respective companies. All other brand and product names are or may be trademarks of, and are used to identify products or services of, their respective owners.

The SCO software that accompanies this publication is commercial computer software and, together with any related documentation, is subject to the restrictions on US Government use as set forth below. If this procurement is for a DOD agency, the following DFAR Restricted Rights Legend applies:

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the US Government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of Rights in Technical Data and Computer Software Clause at DFARS 252.227-7013. Contractor/Manufacturer is The Santa Cruz Operation, Inc., 400 Encinal Street, Santa Cruz, CA 95060.

If this procurement is for a civilian government agency, this FAR Restricted Rights Legend applies:

RESTRICTED RIGHTS LEGEND: This computer software is submitted with restricted rights under Government Contract No. _____ (and Subcontract No. _____, if appropriate). It may not be used, reproduced, or disclosed by the Government except as provided in paragraph (g)(3)(i) of FAR Clause 52.227-14 alt III or as otherwise expressly stated in the contract. Contractor/Manufacturer is The Santa Cruz Operation, Inc., 400 Encinal Street, Santa Cruz, CA 95060.

NonStop Cluster Administration

- Introduction
- Single System Image (SSI) and NSC Administration
- Basic Process Administration
- Filesystem Administration
- Network Administration
- Startup and Shutdown
- Load Leveling
- Failover
- Configuring NSC Behavior
- Troubleshooting Malfunctioning Applications
- Tuning an NSC Kernel

Single System Image (SSI) and NSC Administration

- Overview of SSI in NSC and Current Limitations
- Processes
- File and Device Names
- Networking and CVIP
- Startup and Shutdown

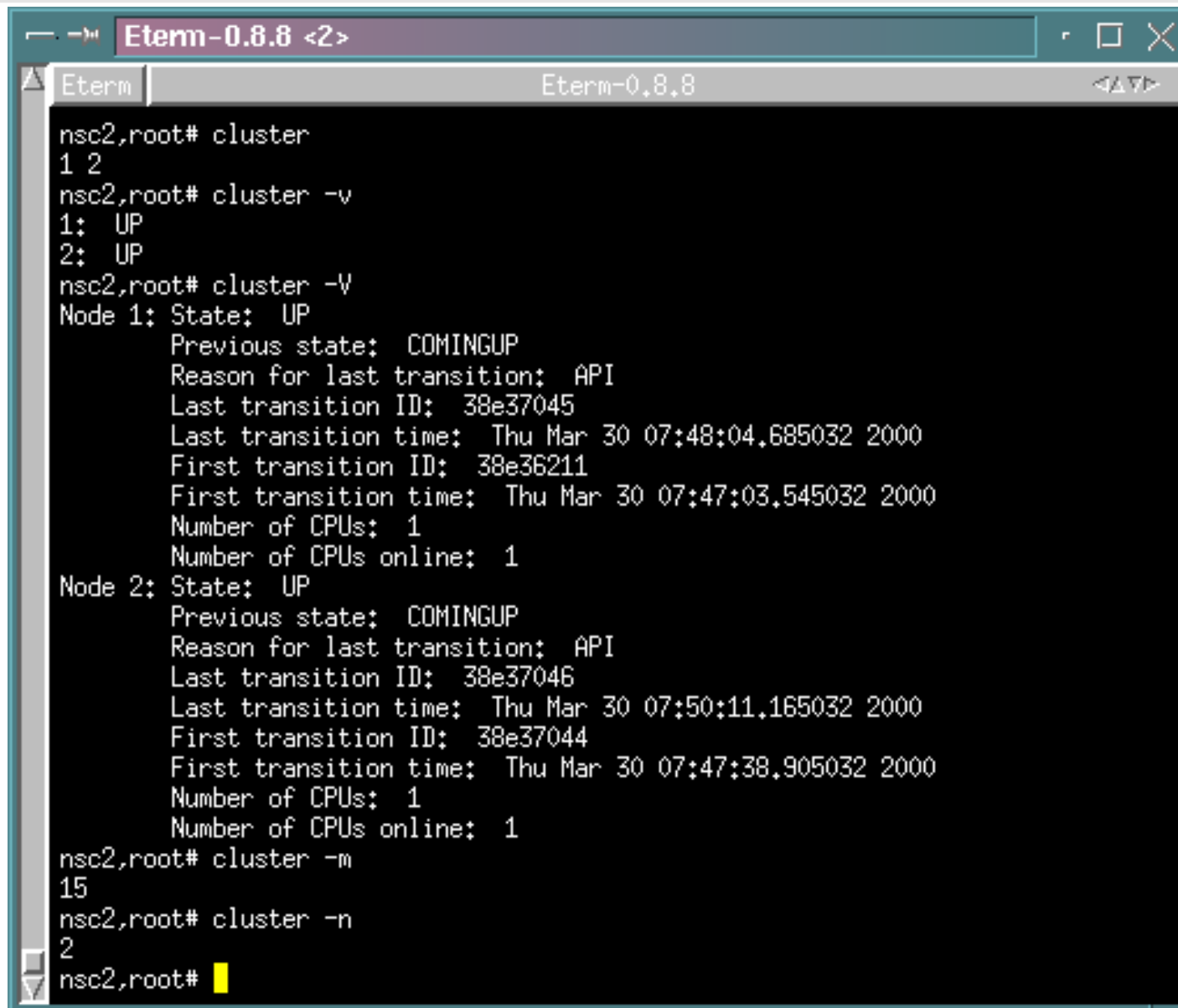
Basic Process Administration

- Overview
- Cluster-specific Process-related Commands
- Multi-Node Execution

Cluster-Specific Process-Related Commands

- **cluster(1)**
- **onall(1)**
- **onnode(1)**
- **node_self(1)**
- **clusternode_num(1)**
- **where(1)**
- **fast(1)**
- **fastnode(1)**
- **kill3(1)**
- **migrate(1)**

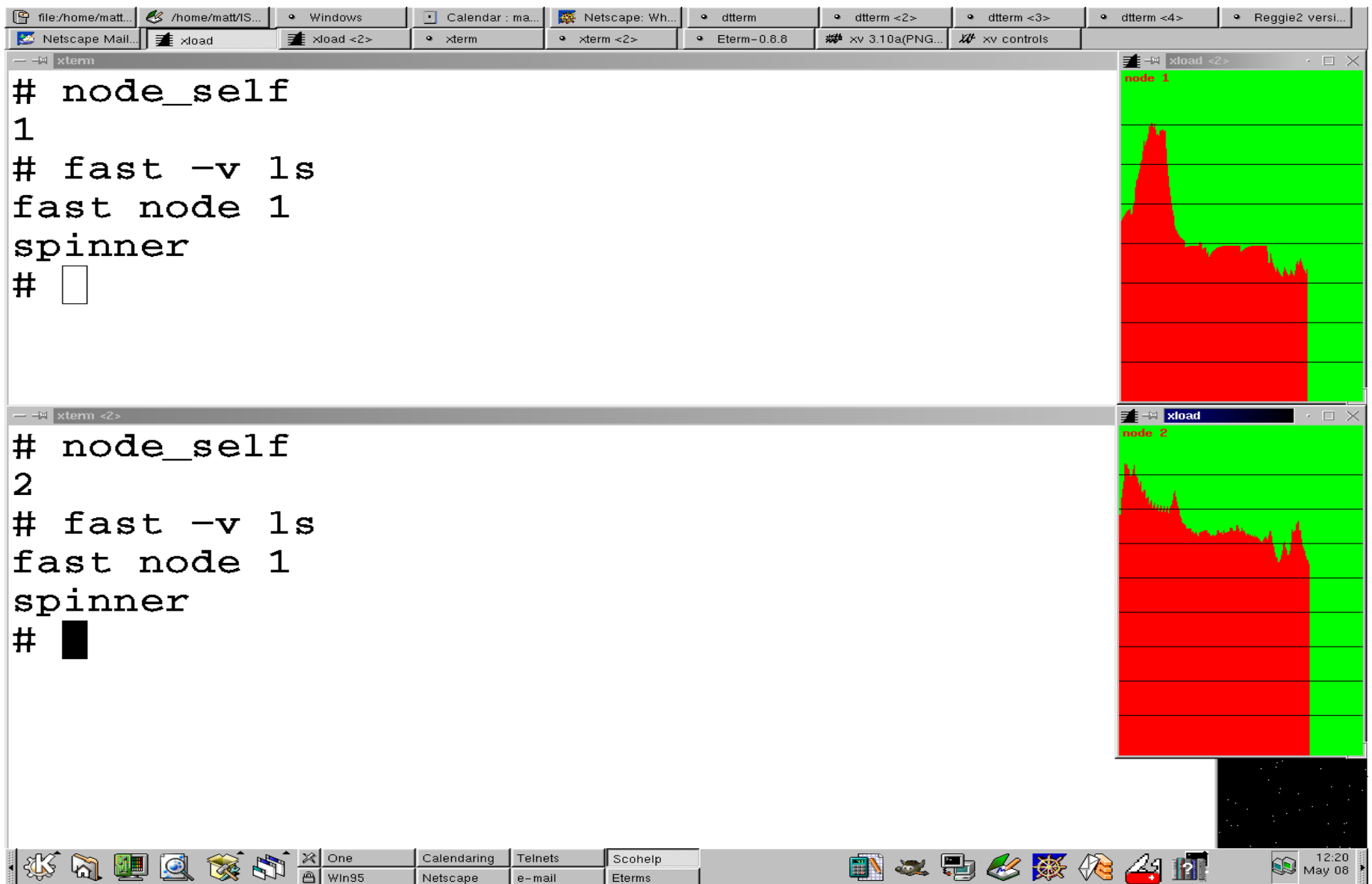
The **cluster(1)** Command



The screenshot shows a terminal window titled "Eterm-0.8.8 <2>". Inside the terminal, the user is at the prompt "nsc2,root#" and has executed several commands. The first command is "cluster", which outputs "1 2". The second command is "cluster -v", which outputs "1: UP" and "2: UP". The third command is "cluster -V", which provides detailed status information for two nodes. Node 1 is in the "UP" state, having previously been "COMINGUP" due to an "API" reason. It shows the last and first transition IDs and times, and indicates 1 CPU is online. Node 2 is also in the "UP" state, with similar transition details and 1 CPU online. The fourth command is "cluster -m", which outputs "15". The fifth command is "cluster -n", which outputs "2". The prompt "nsc2,root#" is followed by a yellow cursor.

```
nsc2,root# cluster
1 2
nsc2,root# cluster -v
1: UP
2: UP
nsc2,root# cluster -V
Node 1: State: UP
        Previous state: COMINGUP
        Reason for last transition: API
        Last transition ID: 38e37045
        Last transition time: Thu Mar 30 07:48:04.685032 2000
        First transition ID: 38e36211
        First transition time: Thu Mar 30 07:47:03.545032 2000
        Number of CPUs: 1
        Number of CPUs online: 1
Node 2: State: UP
        Previous state: COMINGUP
        Reason for last transition: API
        Last transition ID: 38e37046
        Last transition time: Thu Mar 30 07:50:11.165032 2000
        First transition ID: 38e37044
        First transition time: Thu Mar 30 07:47:38.905032 2000
        Number of CPUs: 1
        Number of CPUs online: 1
nsc2,root# cluster -m
15
nsc2,root# cluster -n
2
nsc2,root#
```

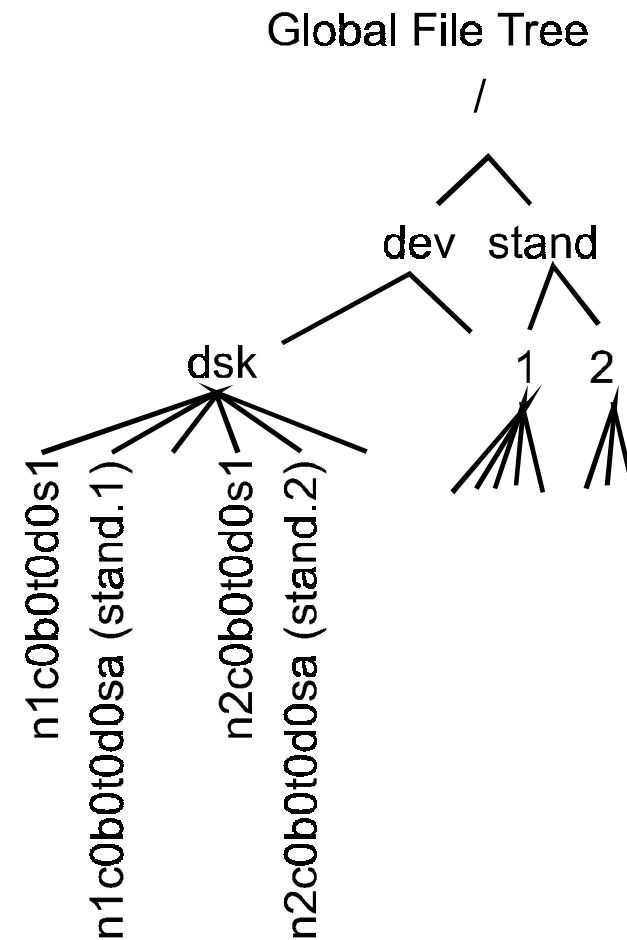
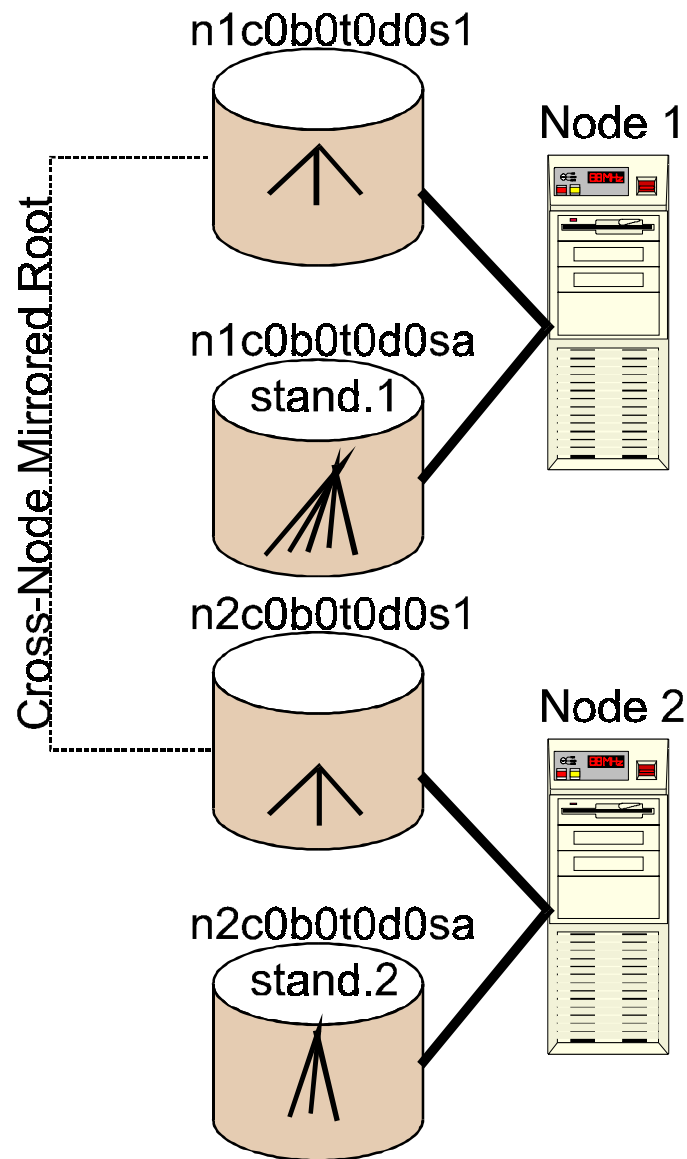
The **fast(1)** Command



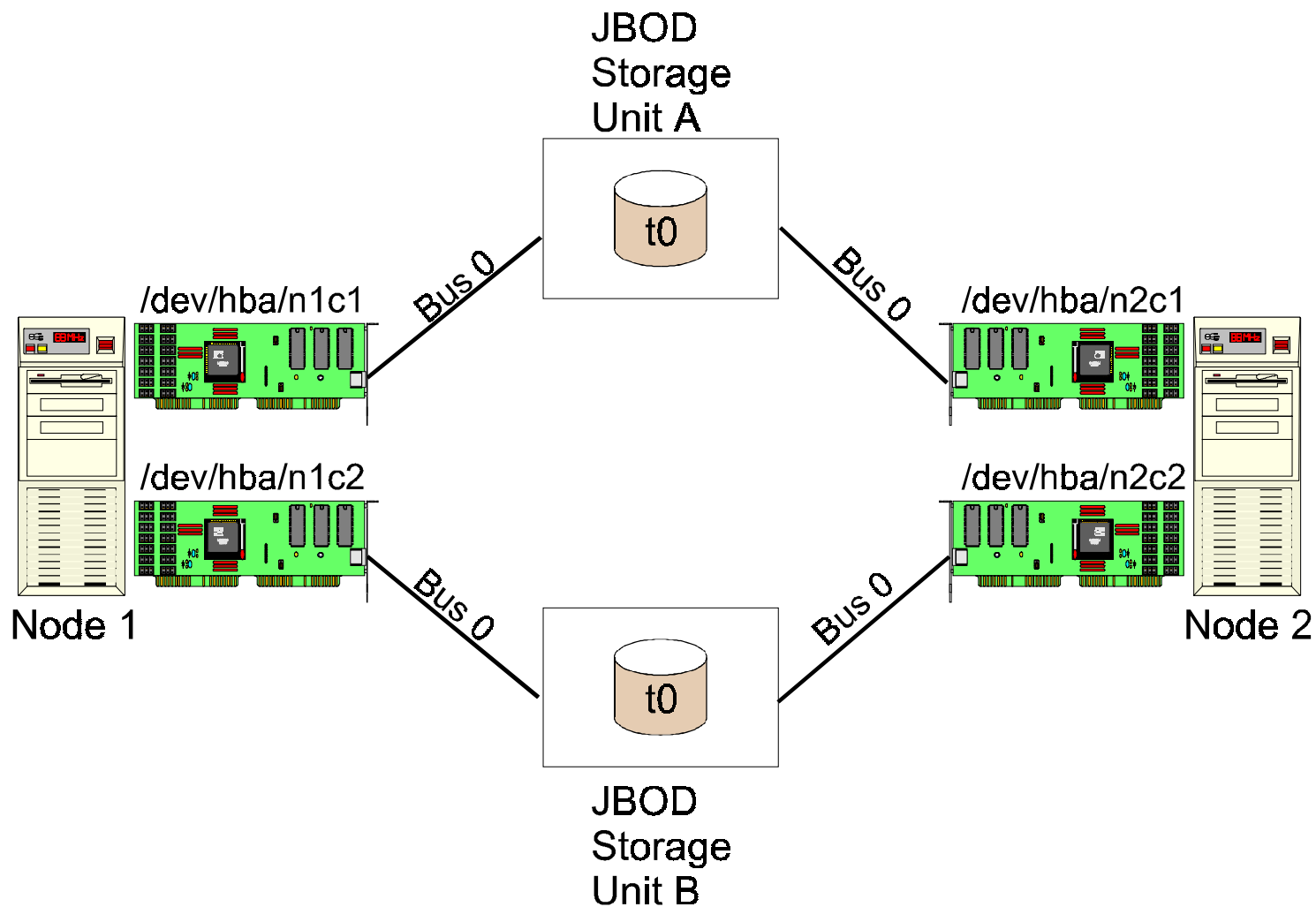
Filesystem Administration

- Overview of Files, Filesystems, and Device Nodes in NSC
- Filesystem Administration In NSC

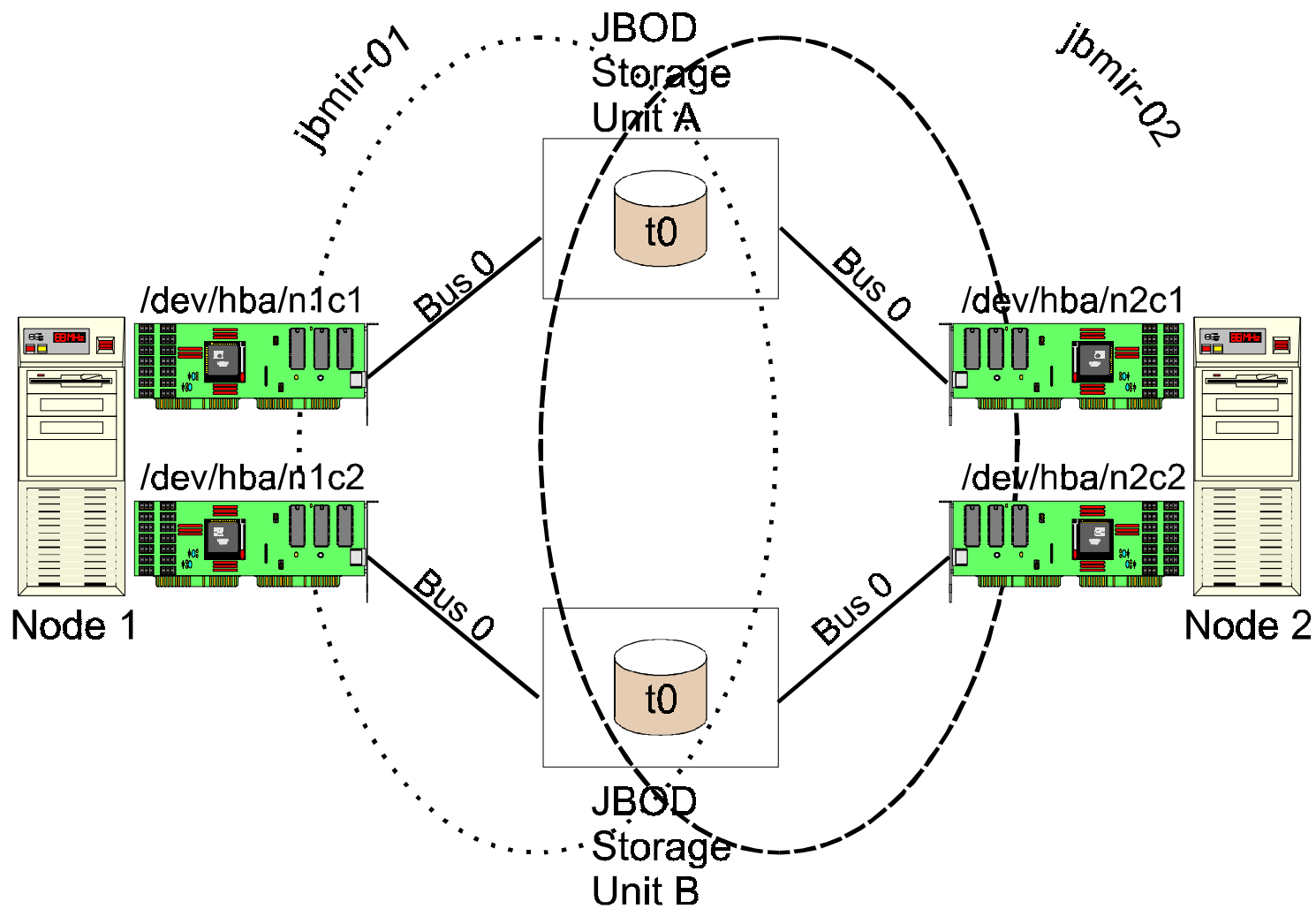
Global File Tree In NSC



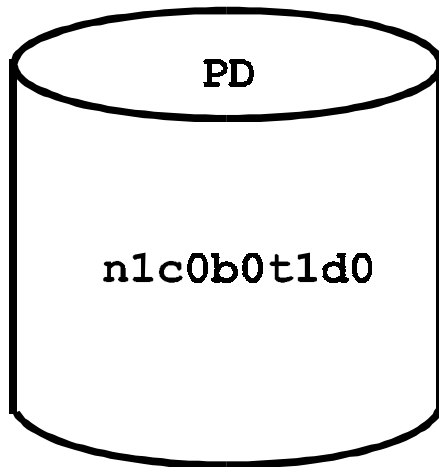
A JBOD Configuration



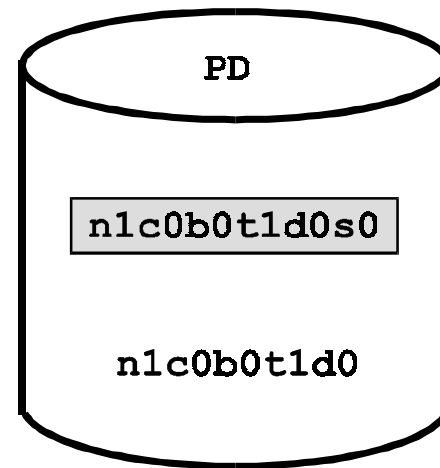
JBOD Mirrors



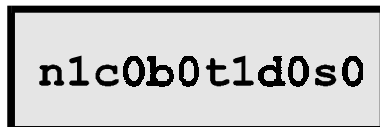
Physical Disk Characteristics



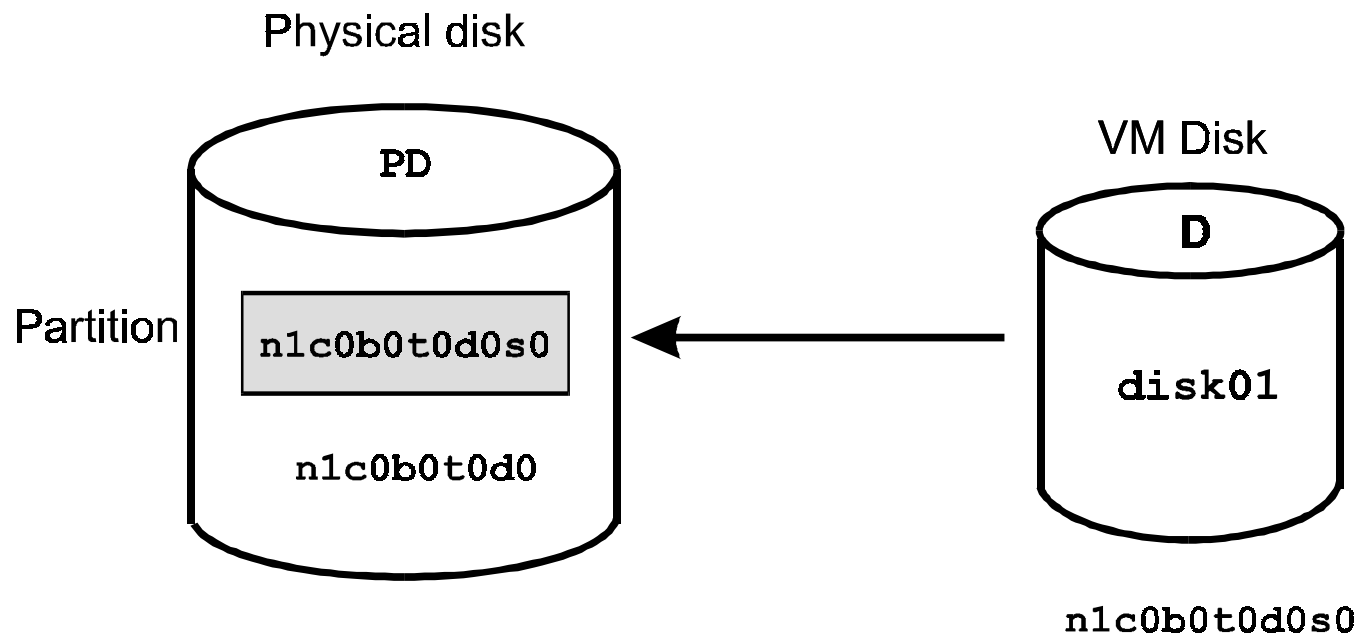
Physical disk with one partition



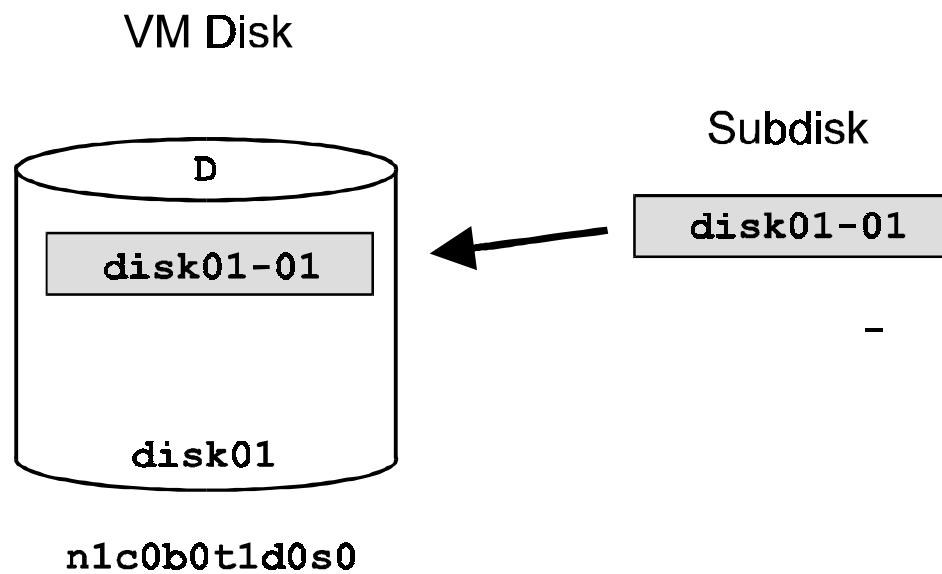
Partition



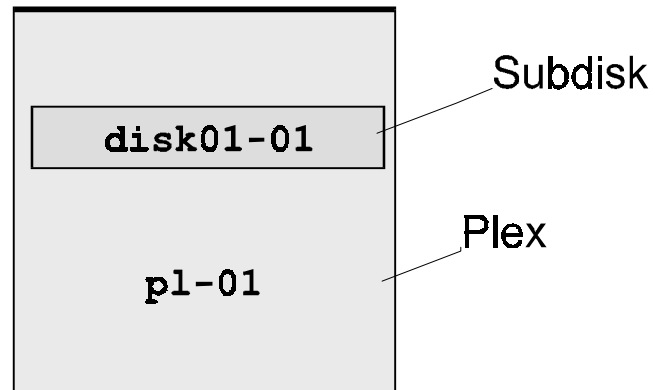
Virtual Disks



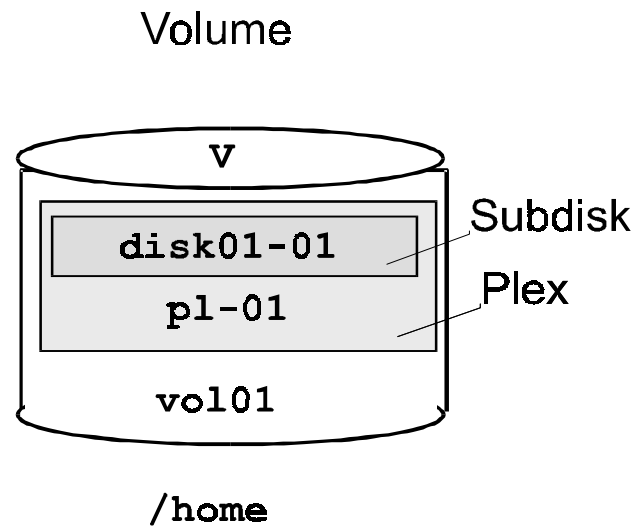
VM Disk with one Subdisk



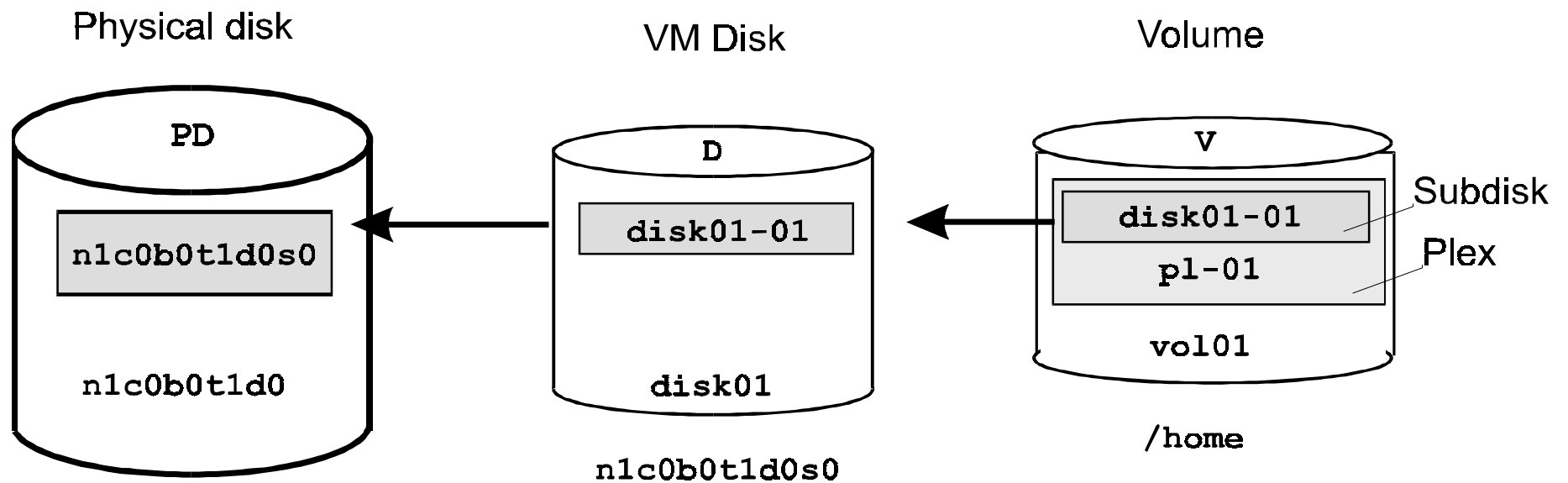
Example of a Plex Composed of One Subdisk



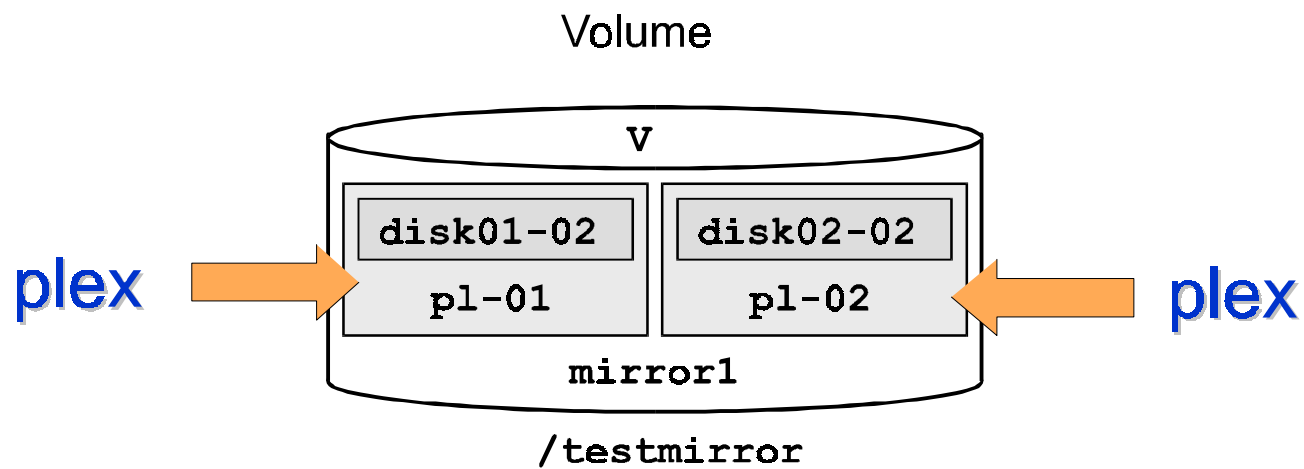
Volume with One Plex



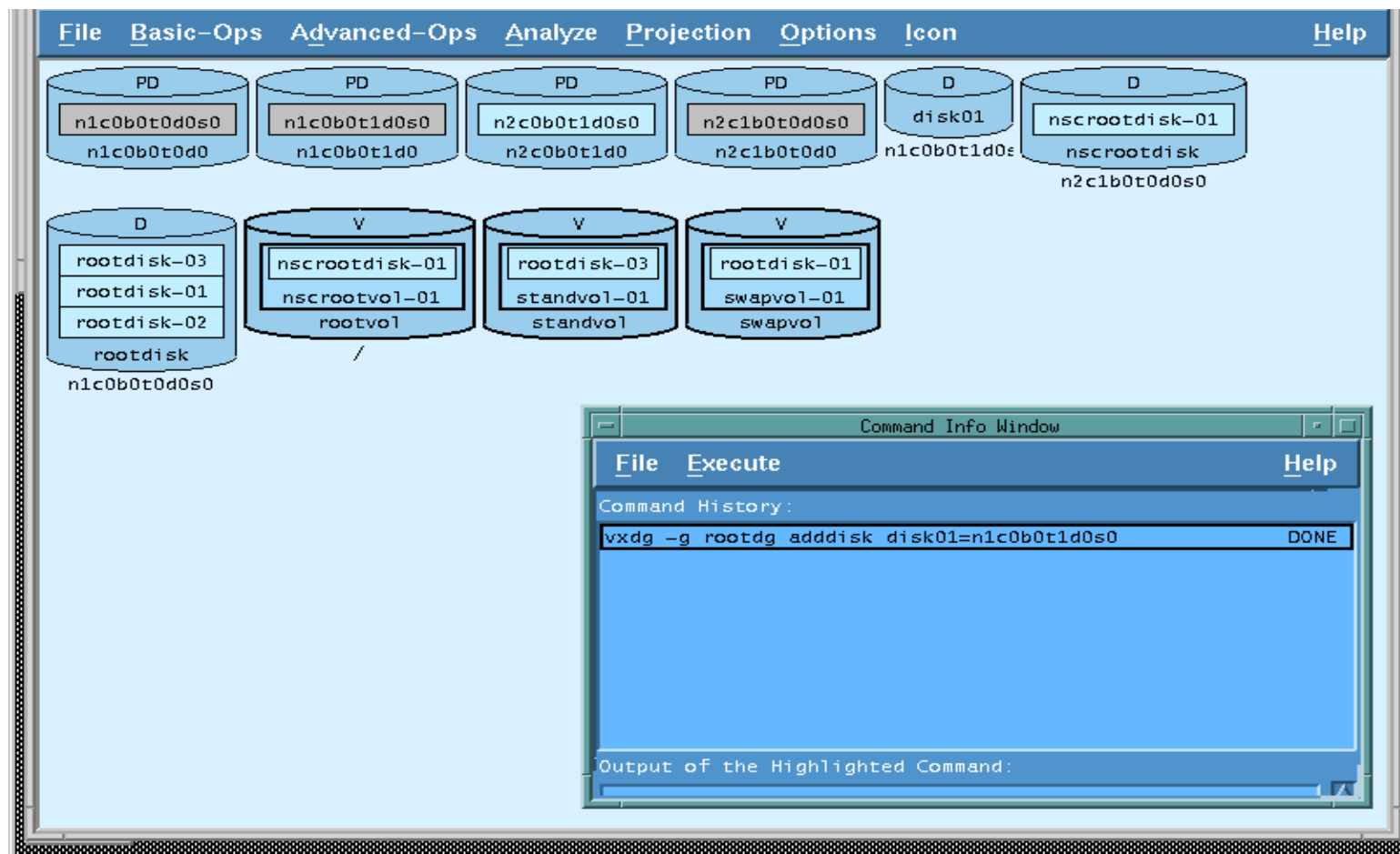
Relationship Between Objects



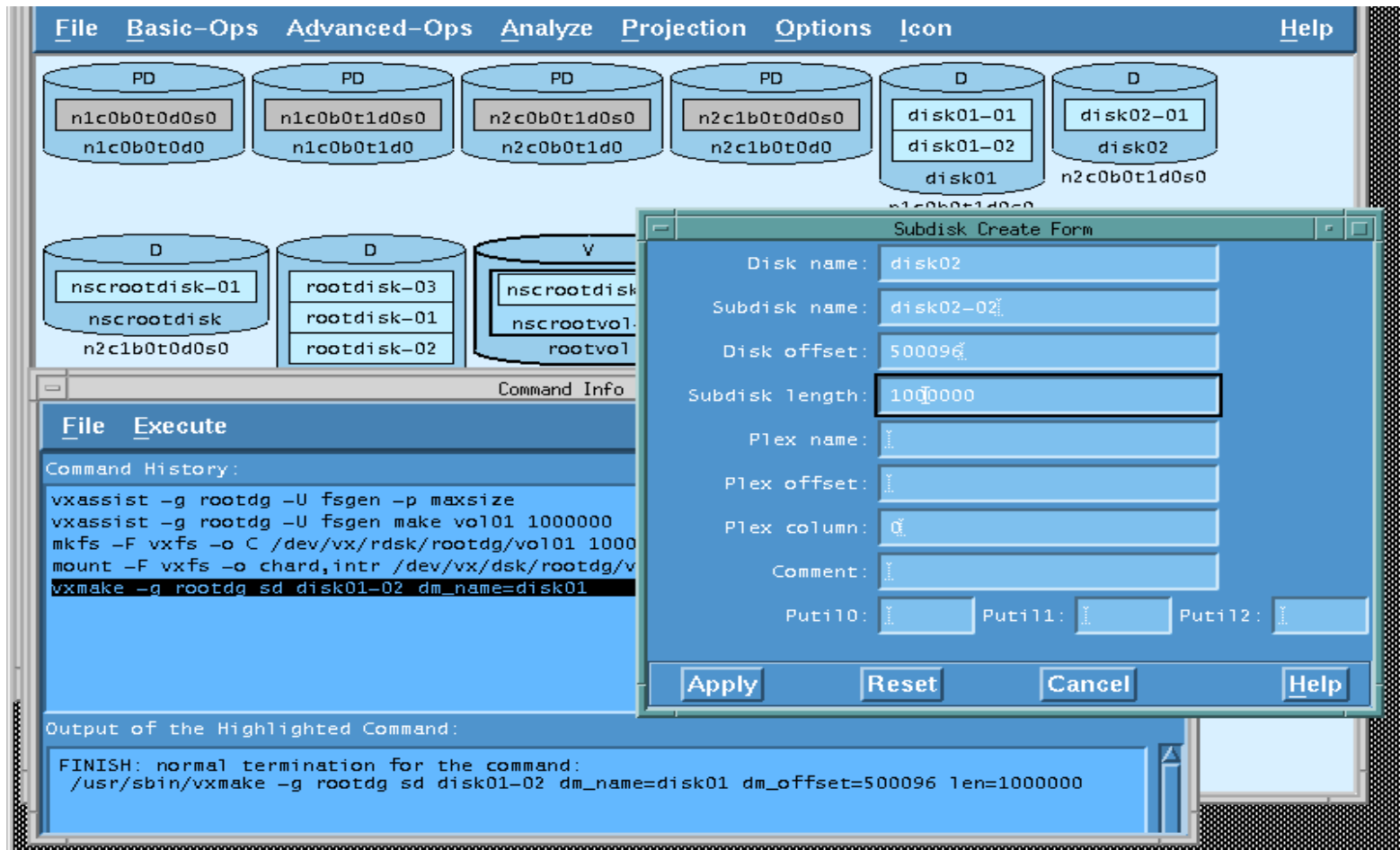
Veritas Disk Mirrors



Adding a VM Disk



Creating a Subdisk



Creating a Plex

The screenshot displays the Filesystem15 GUI with a menu bar (File, Basic-Ops, Advanced-Ops, Analyze, Projection, Options, Icon, Help) and a main workspace containing various disk and volume icons. The icons are organized into two rows. The top row contains six icons: four labeled 'PD' and two labeled 'D'. The bottom row contains six icons: two labeled 'D' and four labeled 'V'. The 'D' icons are labeled 'disk01-01', 'disk01-02', 'disk02-01', and 'disk02-02'. The 'V' icons are labeled 'nscrootdisk-01', 'rootdisk-03', 'rootdisk-01', 'rootdisk-02', 'rootdisk', 'nscrootvol-01', 'standvol-01', 'standvol', 'swapvol-01', 'swapvol', 'disk01-01', 'disk02-01', 'vol101-01', and 'vol101'. The 'disk02-02' icon is highlighted in green. Below the icons, there is a 'Command Info Window' with a 'File' menu and an 'Execute' button. The 'Command History' section shows a list of commands and their status (DONE). The 'Output of the Highlighted Command' section shows the output of the command: `/usr/sbin/vxsd -g rootdg assoc p1-01 disk01-02`.

Filesystem15

Command Info Window

File Execute

Command History:

```
vxassist -g rootdg -U fsgen -p maxsize DONE
vxassist -g rootdg -U fsgen make vol01 1000000 DONE
mkfs -F vxfs -o C /dev/vx/rdisk/rootdg/vol01 1000000 DONE
mount -F vxfs -o chard,intr /dev/vx/dsk/rootdg/vol01 DONE
vxmake -g rootdg sd disk01-02 dm_name=disk01 DONE
vxmake -g rootdg sd disk02-02 dm_name=disk02 DONE
vxmake -g rootdg plex p1-01 DONE
vxmake -g rootdg plex p1-02 DONE
vxsd -g rootdg assoc p1-01 disk01-02 DONE
```

Output of the Highlighted Command:

```
FINISH: normal termination for the command:
/usr/sbin/vxsd -g rootdg assoc p1-01 disk01-02
```

Creating a Volume

The screenshot displays the Filesystem16 graphical user interface. The top menu bar includes: File, Basic-Ops, Advanced-Ops, Analyze, Projection, Options, Icon, and Help. The main workspace shows a hierarchical disk layout:

- PD (Physical Disk) Level:** Four disks labeled n1c0b0t0d0s0, n1c0b0t1d0s0, n2c0b0t1d0s0, and n2c1b0t0d0s0.
- D (Disk) Level:**
 - disk01-01, disk01-02 (grouped as disk01, n1c0b0t1d0s0)
 - disk02-01, disk02-02 (grouped as disk02, n2c0b0t1d0s0)
 - nscrootdisk-01, nscrootdisk (grouped as nscrootvol-01, n2c1b0t0d0s0)
 - rootdisk-03, rootdisk-01, rootdisk-02 (grouped as rootdisk, n1c0b0t0d0s0)
- V (Volume) Level:**
 - mirror1
 - nscrootvol-01
 - standvol-01
 - swapvol-01
 - A volume group containing disk01-01, disk02-01, vo101-01, vo101, and /test_stripe.
- Planes (p1-01, p1-02):** Located below the volume group, associated with disk01-02 and disk02-02.

A "Command Info Window" is open in the foreground, showing the following command history:

```

vxassist -g rootdg -U fsgen make vol01 1000000 DONE
mkfs -F vxfs -o C /dev/vx/rdisk/rootdg/vol01 1000000 DONE
mount -F vxfs -o chard,intr /dev/vx/dsk/rootdg/vol01 DONE
vxmake -g rootdg sd disk01-02 dm_name=disk01 DONE
vxmake -g rootdg sd disk02-02 dm_name=disk02 DONE
vxmake -g rootdg plex p1-01 DONE
vxmake -g rootdg plex p1-02 DONE
vxsd -g rootdg assoc p1-01 disk01-02 DONE
vxsd -g rootdg assoc p1-02 disk02-02 DONE
vxmake -U fsgen -g rootdg vol mirror1 DONE
  
```

Below the command history, the "Output of the Highlighted Command:" is shown:

```

vxvm:vxmake: WARNING: Volume mirror1: length set to zero
FINISH: normal termination for the command:
/usr/sbin/vxmake -U fsgen -g rootdg vol mirror1 read_pol=SELECT user:
  
```

Adding a Plex to a Volume

The screenshot shows the VxFS File System Manager interface. The main area displays various storage objects:

- PD (Physical Disk) Objects:**
 - disk01-01, disk01-02 (n1c0b0t0d0s0, n1c0b0t0d0)
 - disk02-01, disk02-02 (n2c0b0t1d0s0, n2c0b0t1d0)
 - disk01-02, p1-01 (n2c1b0t0d0s0, n2c1b0t0d0)
 - rootdisk-03, rootdisk-01, rootdisk-02 (n1c0b0t1d0s0, n1c0b0t1d0)
 - nsrootdisk-01, nsrootdisk (n2c1b0t0d0s0, n2c1b0t0d0)
- D (Disk) Objects:**
 - disk01-01, disk01-02 (n1c0b0t1d0s0)
 - disk02-01, disk02-02 (n2c0b0t1d0s0)
- V (Volume) Objects:**
 - vol101-01, vol101 (/test_stripe)
 - mirror1 (p1-01)
 - rootvol (rootdisk-01, nsrootvol-01)
 - standvol (standvol-01)
 - swapvol (swapvol-01)

The **Command Info Window** is open, showing the following command history:

```

mount -F vxfs -o chard,intr /dev/vx/dsk/rootdg/vol101  DONE
vxmake -g rootdg sd disk01-02 dm_name=disk01          DONE
vxmake -g rootdg sd disk02-02 dm_name=disk02          DONE
vxmake -g rootdg plex p1-01                            DONE
vxmake -g rootdg plex p1-02                            DONE
vxsd -g rootdg assoc p1-01 disk01-02                  DONE
vxsd -g rootdg assoc p1-02 disk02-02                  DONE
vxmake -U fsgen -g rootdg vol mirror1                 DONE
vxplex -g rootdg att mirror1 p1-01                    DONE
vxvol -g rootdg set len=1000000 mirror1                DONE
  
```

The output of the highlighted command is:

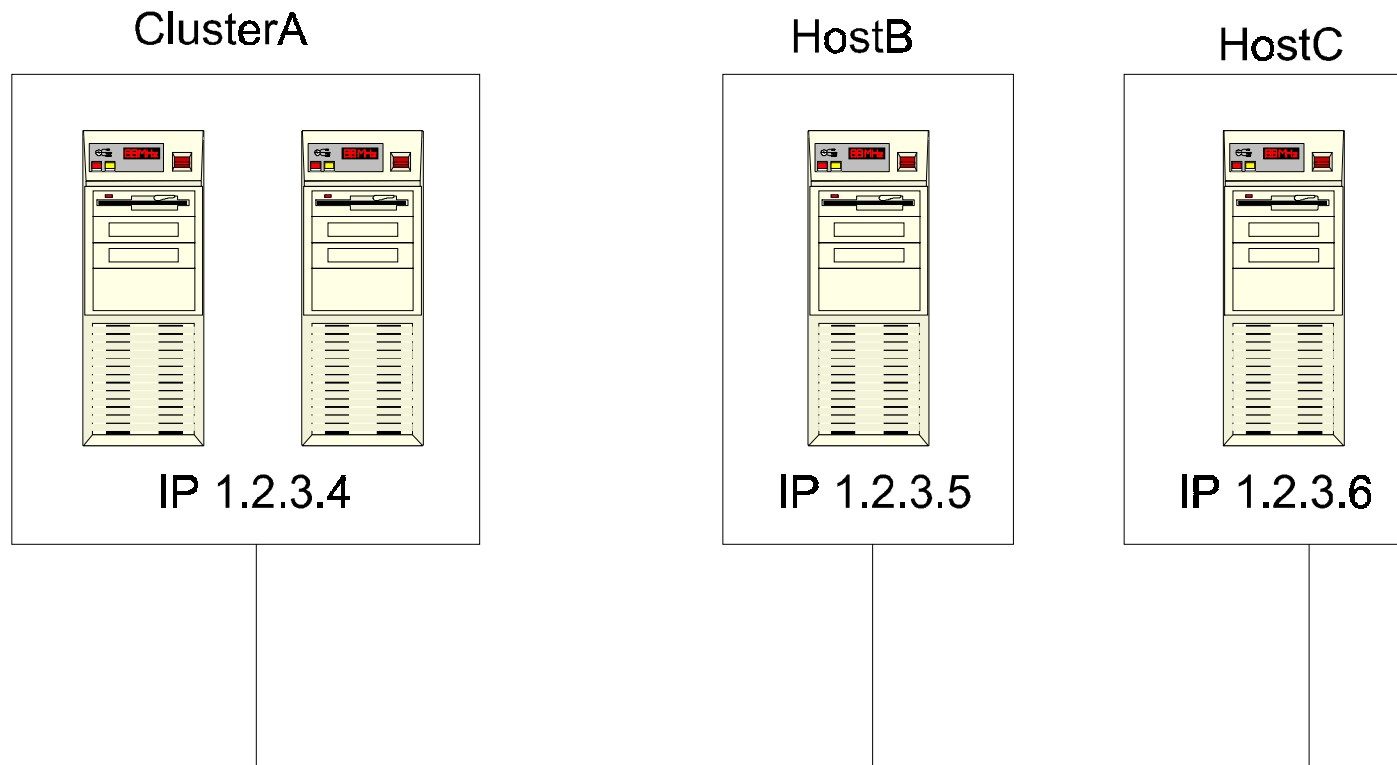
```

FINISH: normal termination for the command:
/usr/sbin/vxvol -g rootdg set len=1000000 mirror1
  
```

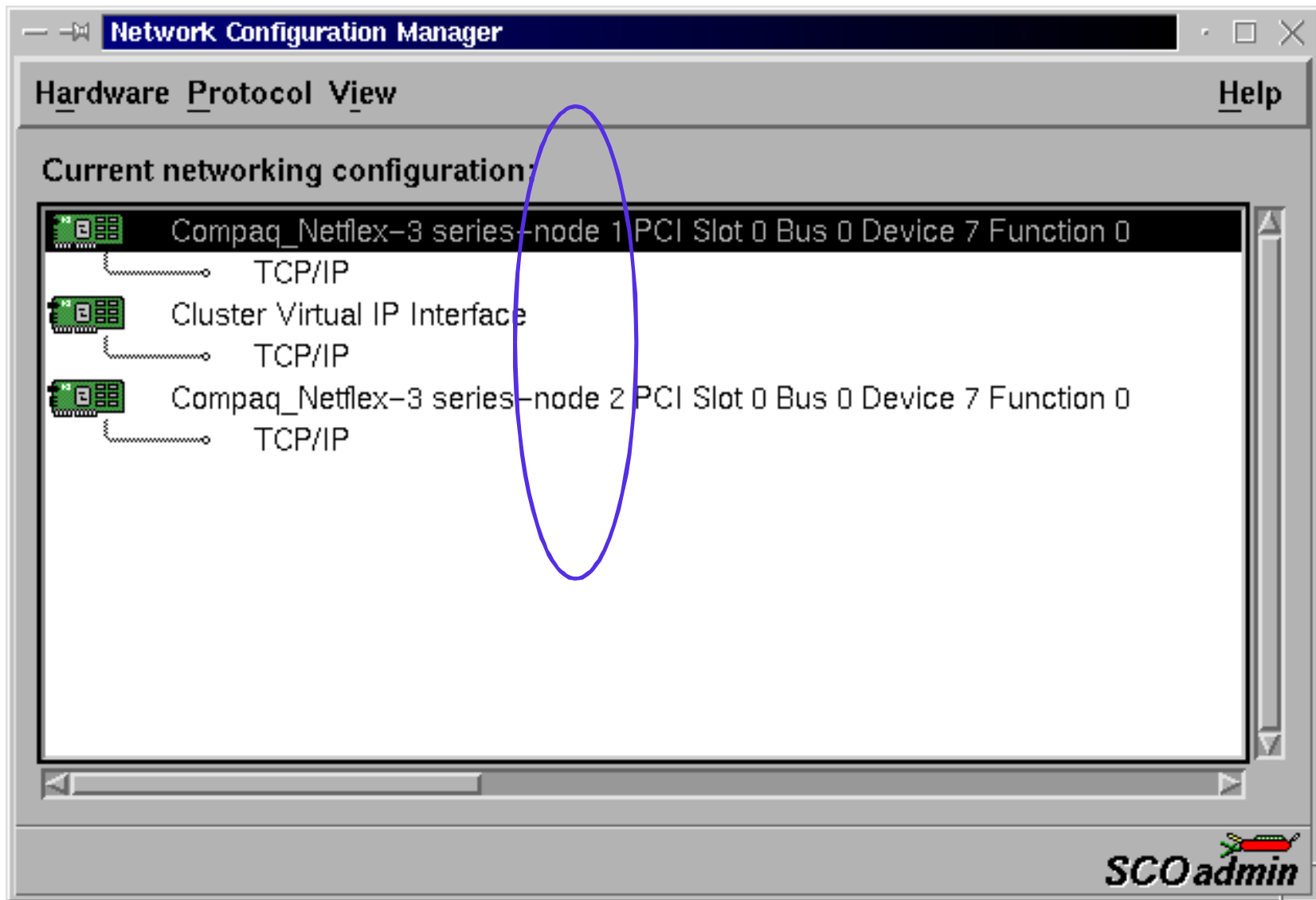
Network Administration

- Overview of Networking in NSC
- Adding NICs and PIPs
- The CVIP
- IP-Aliasing in NSC
- Packet Routing in NSC
- Running Multiple Server Daemons
- PPP

SSI And Networking



The **netcfg**(1M) Command In NSC

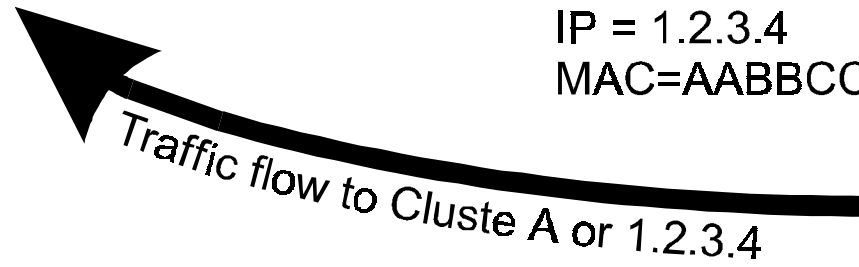


Initial CVIP Mapping

ClusterA Node 1
IP = 1.2.3.41
MAC=AABBCCDD

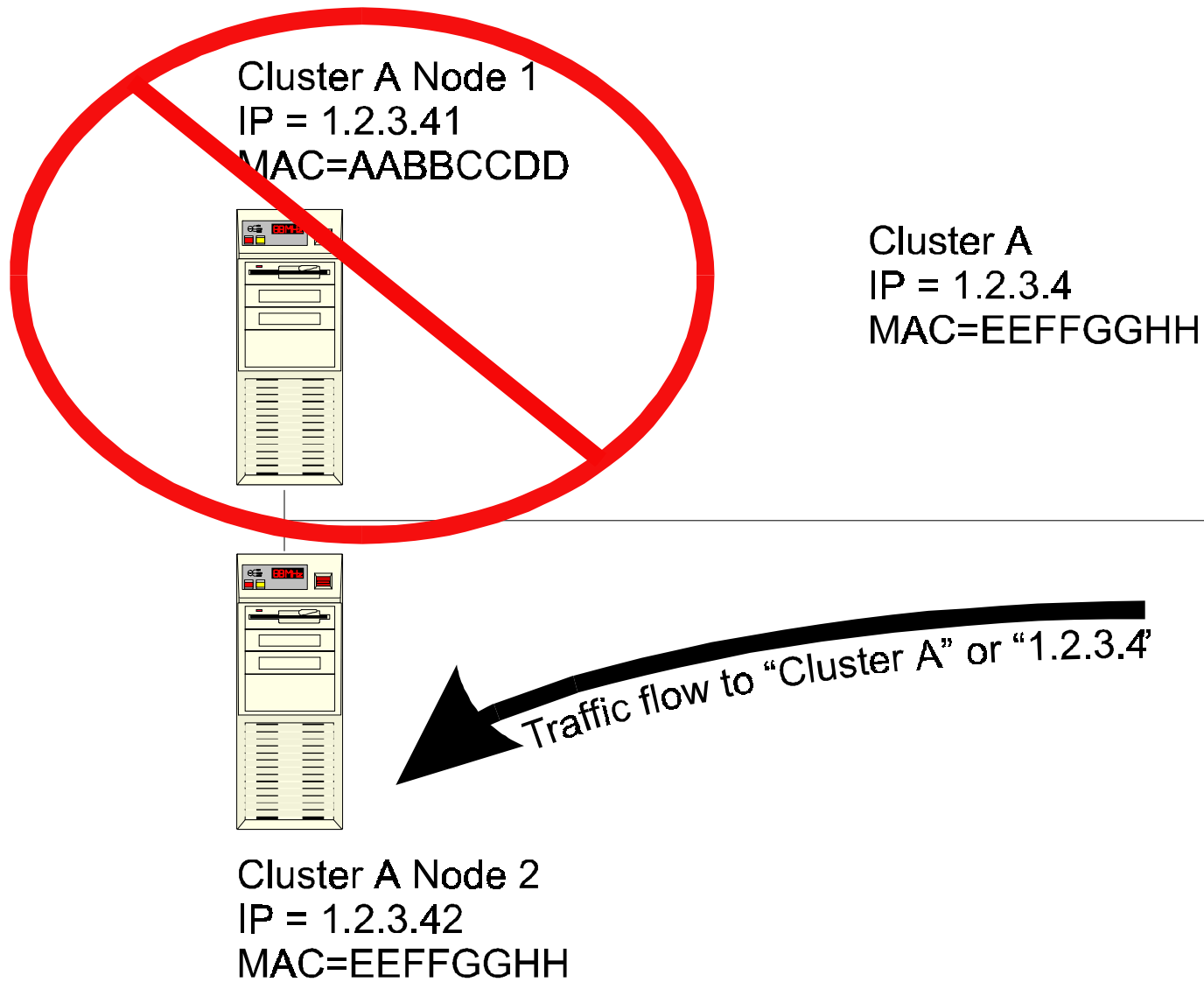


ClusterA
IP = 1.2.3.4
MAC=AABBCCDD

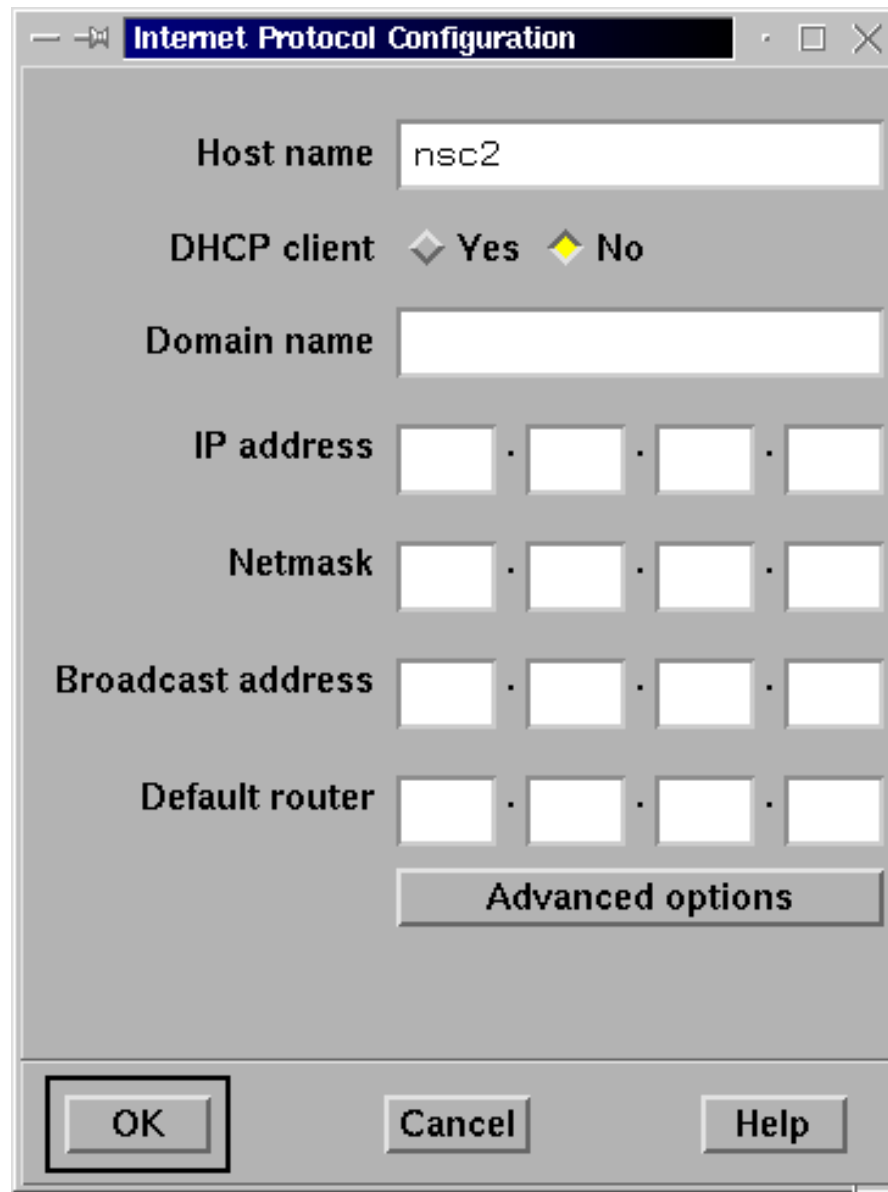


ClusterA Node 2
IP = 1.2.3.42
MAC=EEFFGGHH

CVIP Mapping After Node 1 Failure



Adding A CVIP



The image shows a screenshot of a Windows-style dialog box titled "Internet Protocol Configuration". The dialog box has a title bar with standard minimize, maximize, and close buttons. The main area contains several fields and options:

- Host name:** A text box containing the value "nsc2".
- DHCP client:** A section with two radio buttons. The "Yes" button is selected (indicated by a diamond), and the "No" button is unselected (indicated by a yellow diamond).
- Domain name:** An empty text box.
- IP address:** Four empty text boxes separated by dots, representing the IP address fields.
- Netmask:** Four empty text boxes separated by dots, representing the netmask fields.
- Broadcast address:** Four empty text boxes separated by dots, representing the broadcast address fields.
- Default router:** Four empty text boxes separated by dots, representing the default router fields.
- Advanced options:** A button located below the IP address fields.

At the bottom of the dialog box, there are three buttons: "OK", "Cancel", and "Help". The "OK" button is highlighted with a black border.

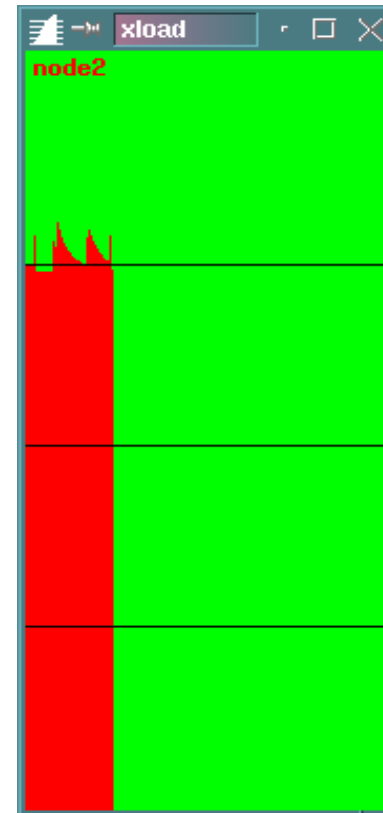
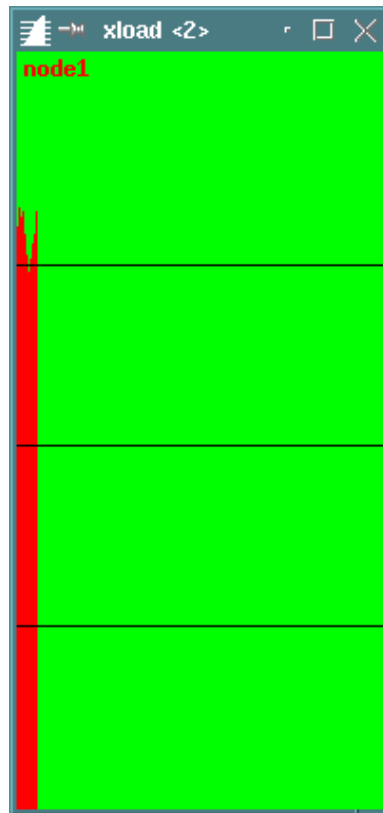
Startup and Shutdown

- Cluster Startup
- Configuring Startup
- Shutdown For The Entire Cluster
- Shutdown and Failure on Individual Nodes

Load Leveling

- Load Leveling Goals
- Load Leveling Concepts
- The Mechanics of Load Leveling
- Configuring Load Leveling
- Starting and Stopping the Load Leveling Daemons
- Overriding Load Leveling
- Load Leveling Domains

A Load Levelled Cluster



The *parameters*(4) File

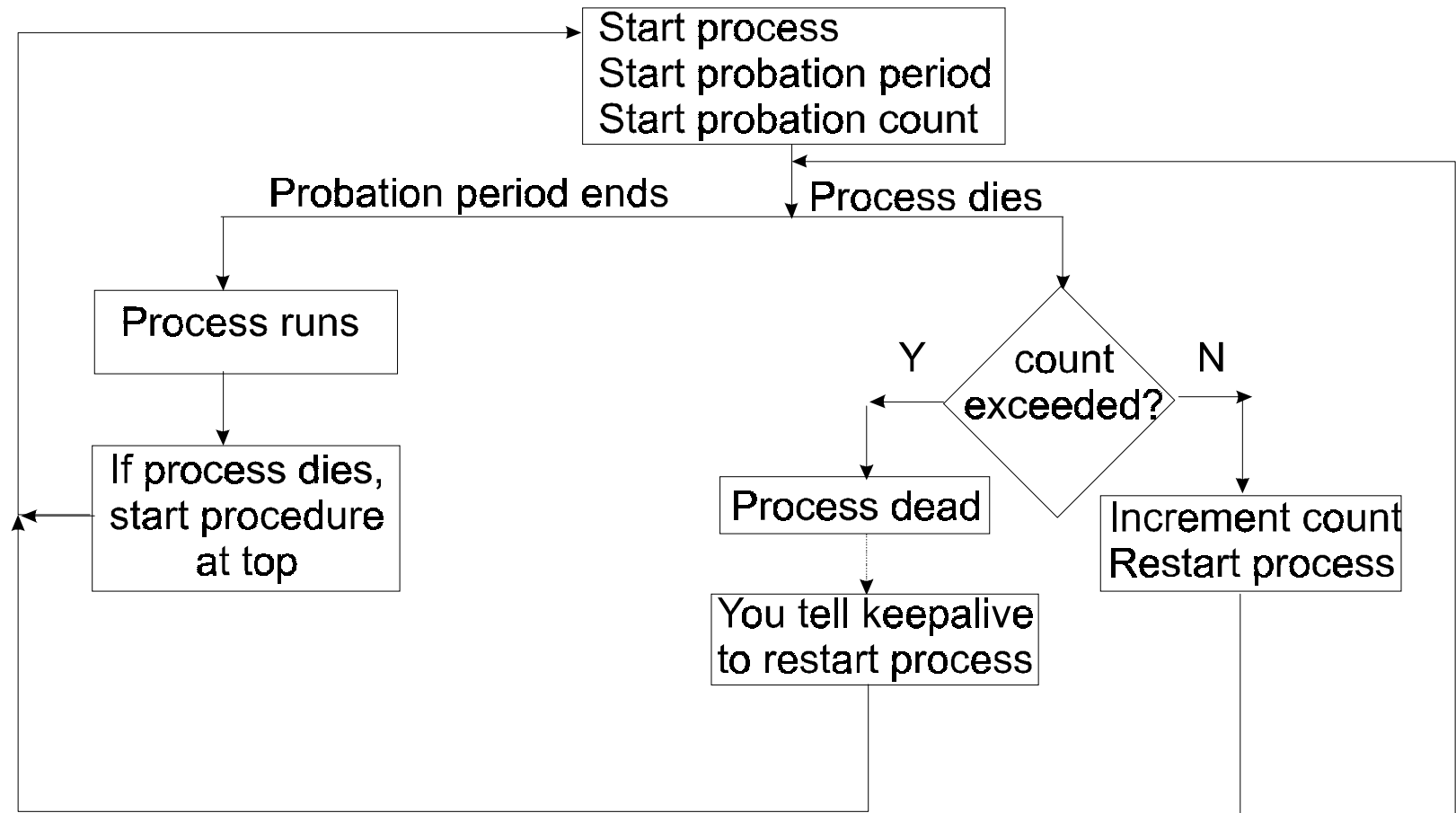
```
#
# This file specifies the parameters used by the load leveler daemon

send_timeout      2.0 # frequency in secs for sending load info.
redispatch_timeout 5.0 # frequency in seconds for re-dispatching procs
first_weight_factor 50 # 5 second weight factor
second_weight_factor 25 # 30 second weight factor
third_weight_factor 25 # 1 minute weight factor
number_vector_elements 8 # number of load vector elements
migrate_inclusion_flag 0 # 0 = exclusion list, 1 = inclusion list
rexec_inclusion_flag 0 # 0 = exclusion list, 1 = inclusion list
minimum_underload 1.0 # node is underloaded if underload <= parameter
minimum_overload 2.0 # node is overloaded if overload >= parameter
minimum_cputime 4.0 # minimum runtime of process for migration
per_process_load 1.0 # load attributed to a single process
nodes_to_use all # specify all nodes
rexec_bias 4.0 # per node amount to bias rexec
```

Failover

- Review of Failover Concepts
- Understanding Failover
- Configuring Failover
- Restarting Databases
- The **NCMS Keepalive Manager**

Process Probation Algorithm



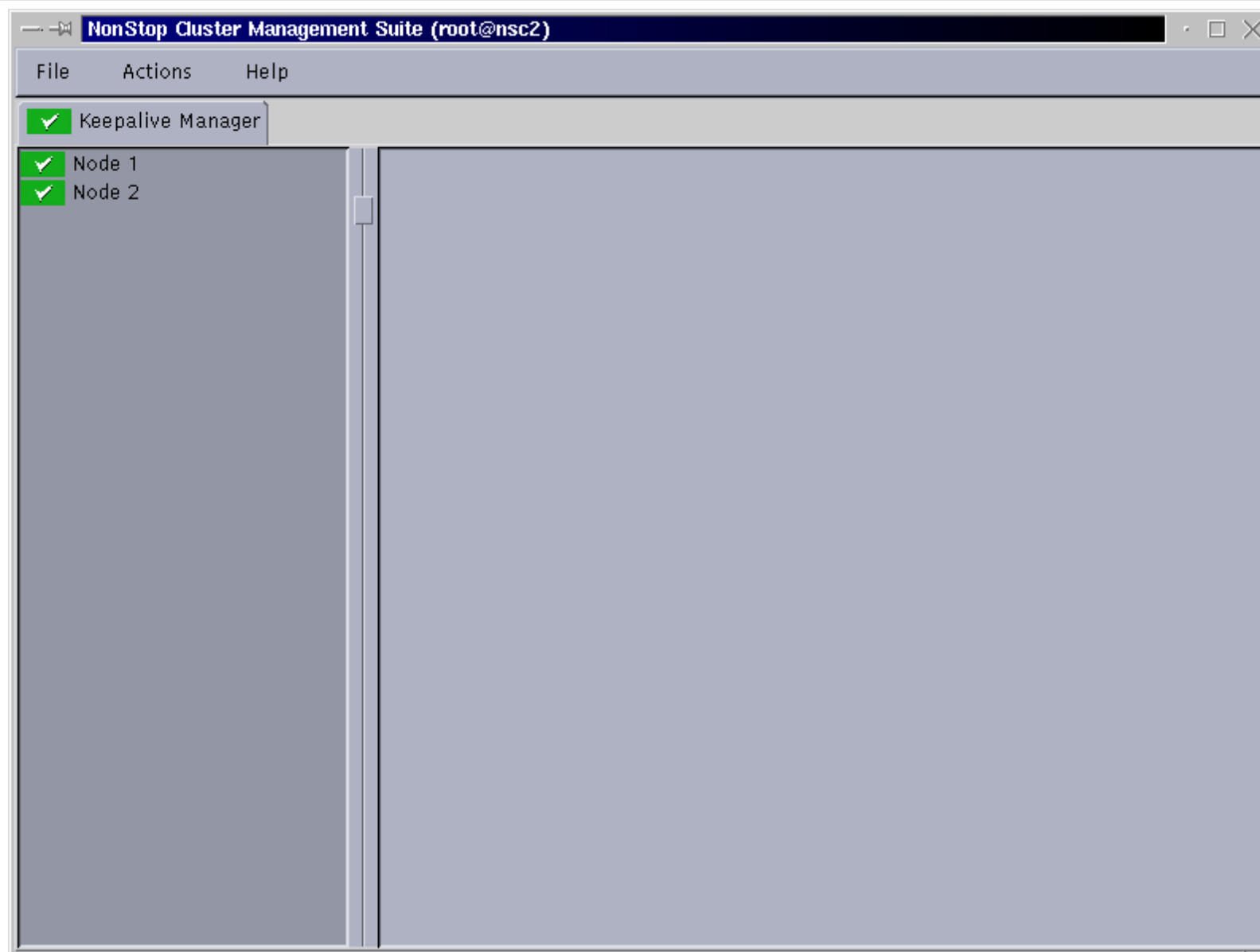
The `/etc/keepalive.d/load_level_d_restart` File

```
#!/sbin/sh
DAEMON="/usr/sbin/load_level_d"
DAEMON_OPTS="-f -k -v"
if [ ! -x $DAEMON ]
then
    echo "load_level_d is disabled"
    exit 0
fi
exec $DAEMON $DAEMON_OPTS
```

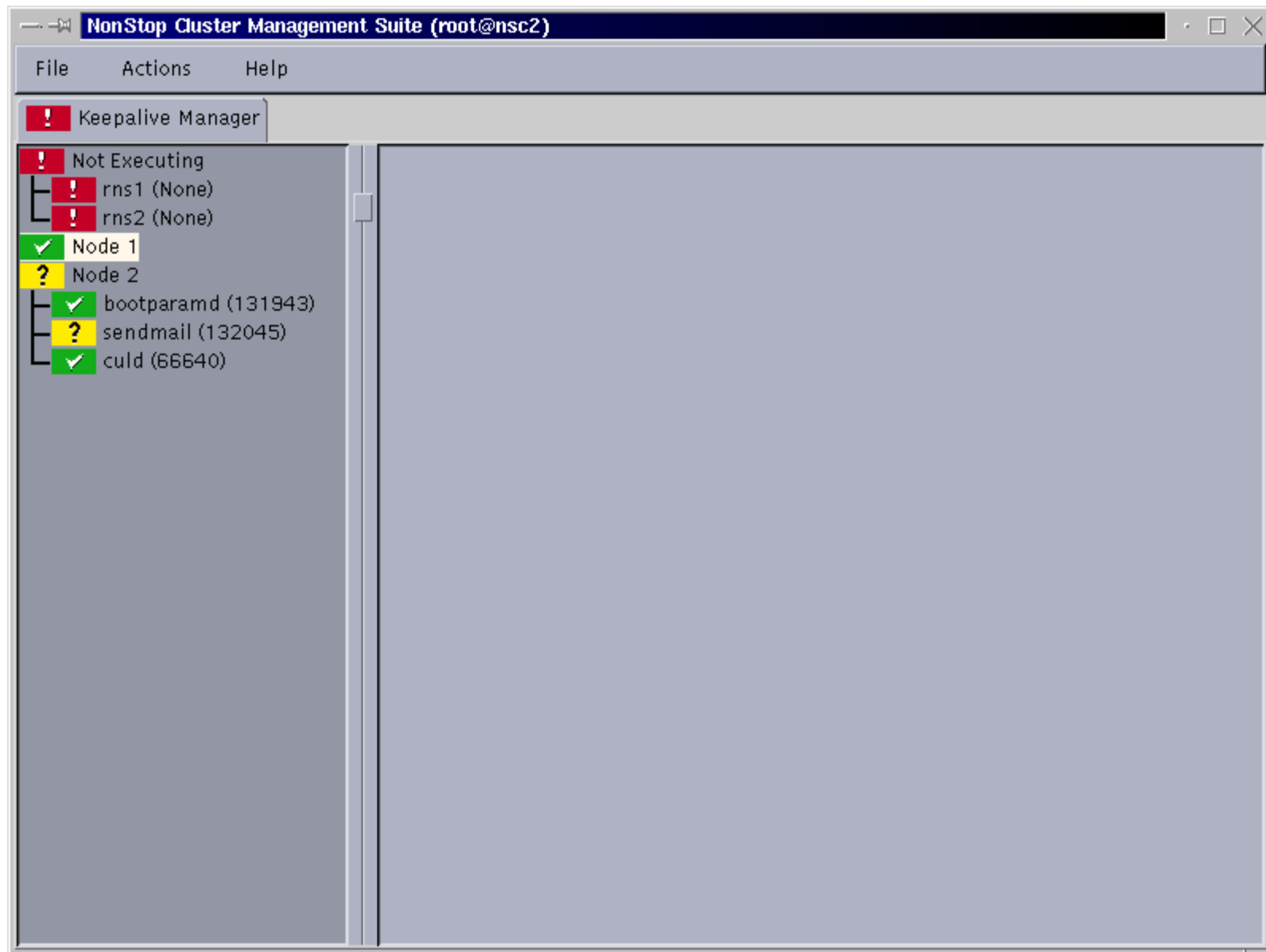
The `/etc/spawndaemon.d/ka_load_leveld` File

```
:/usr/sbin/load_leveld:-f -k -v:60:root:sys::::load_leveld_restart:::::
```

The NCMS Keepalive Manager



Keepalive Manager With Expanded Folders



Process Selection in Keepalive Manager

NonStop Cluster Management Suite (root@nsc2)

File Actions Help

! Keepalive Manager

! Not Executing

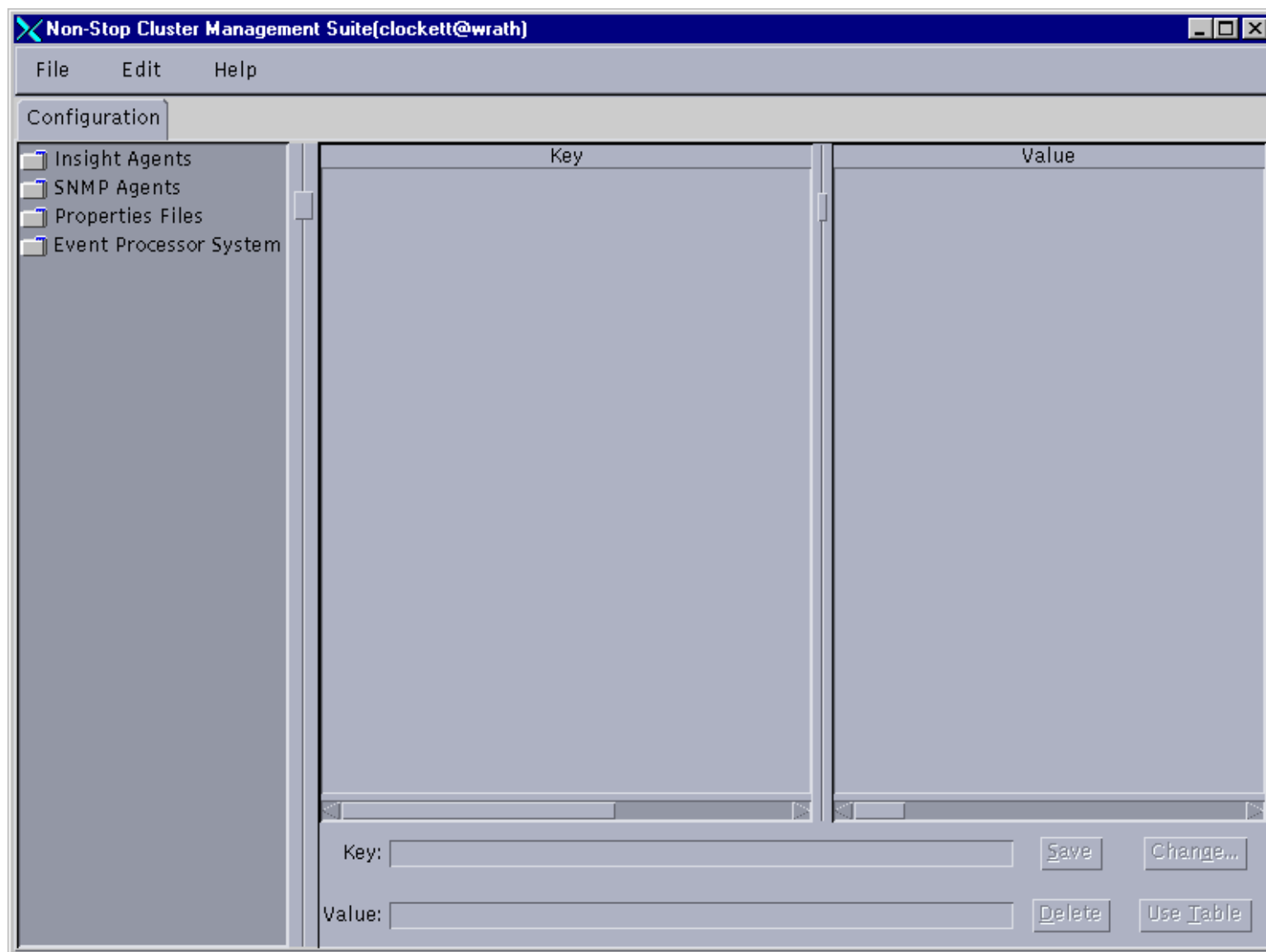
- ! rns1 (None)
- ! rns2 (None)
- ✓ Node 1
- ? Node 2
 - ✓ bootparamd (131943)
 - ? sendmail (132045)
 - ✓ culd (66640)

Attribute Name	Attribute Value
state	ok
pid	131943
node_number	2
full_path_to_process	/usr/lib/nfs/bootparamd
arg_list	
child_of_keepalive	TRUE
daemonization_recovery	TRUE
pinned	TRUE
lastexeced	Mon Apr 3 17:26:20 2000
process_first_died	Never
process_last_died	Never
minrespawn	0
num_errors	0
total_errors	0
max_errors_during_probation	10
probation_period	300
registration_policy	Name
node_selection_policy	ROUND_ROBIN
favored_node	None
backup_nodes	None
rejected_nodes	None
termwait	2
euid	0
egid	3
startup_script	/etc/keepalive.d/bootparamd_restart
shutdown_script	None
process_failure_recovery_script	None
node_failure_recovery_script	None
down_script	None

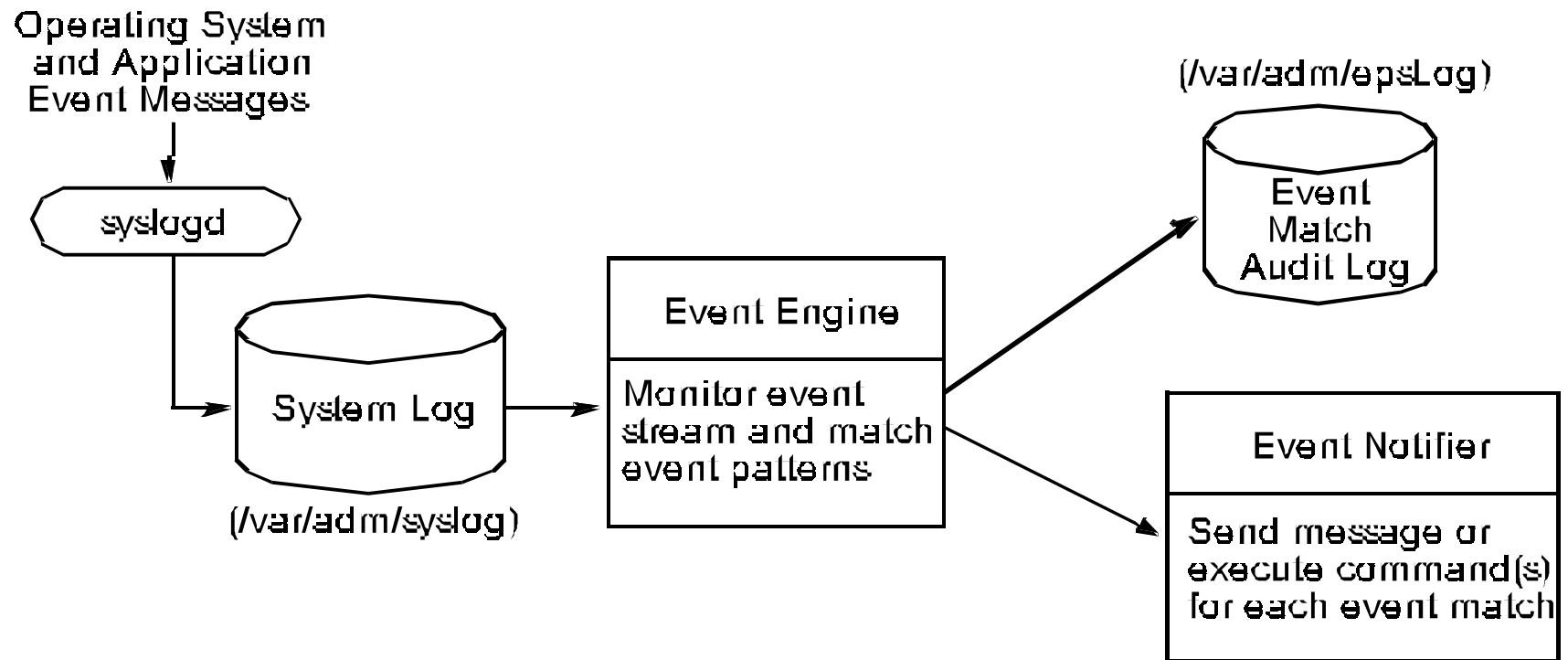
Configuring NSC Behavior

- Overview of NCMS
- EPS
- Other NCMS Configuration
- Other NSC Configuration Files

NCMS Configuration Manager



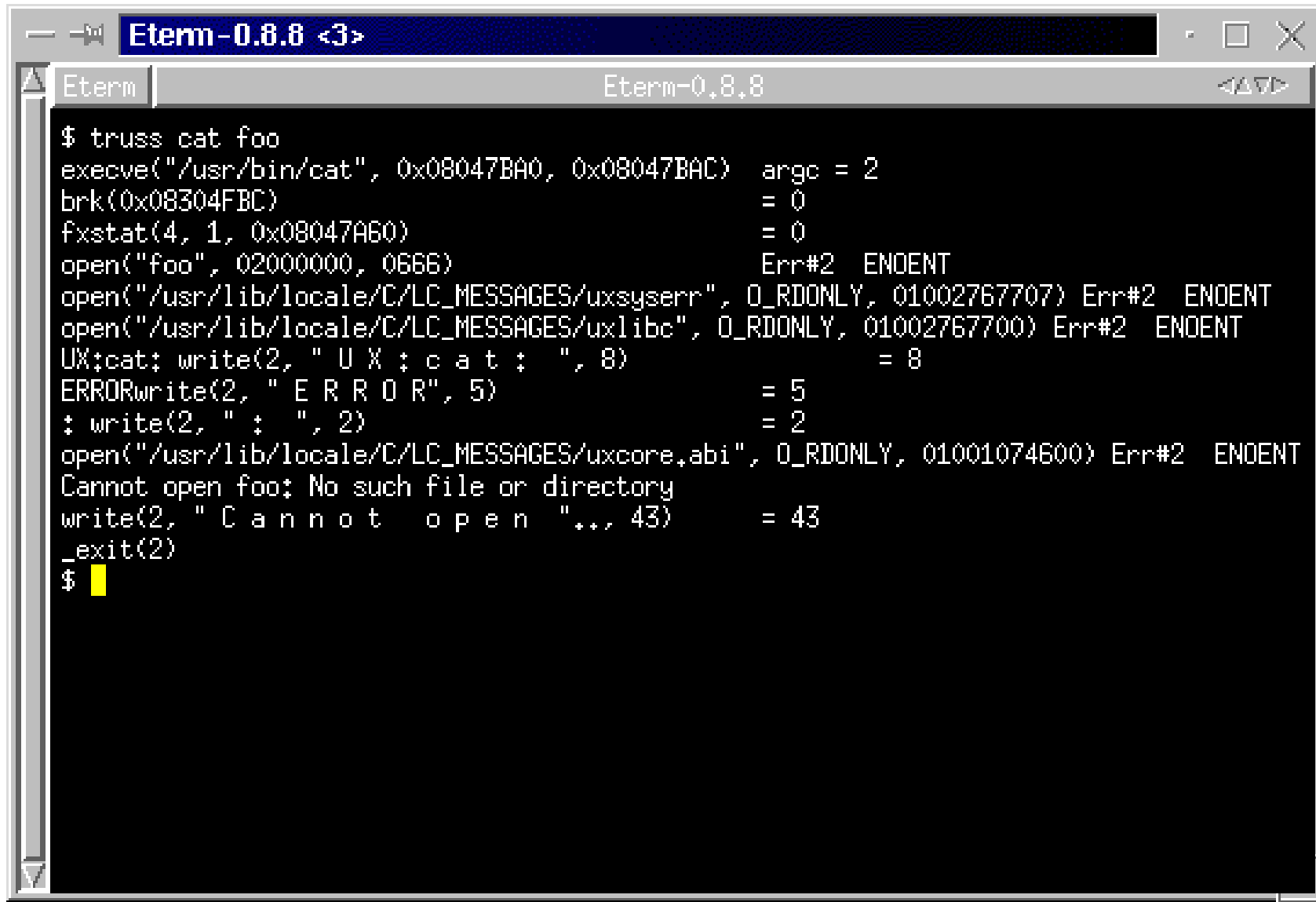
How EPS Works



Troubleshooting Malfunctioning Applications

- Detecting and Solving Typical Problems
- The **truss**(1) Command
- Other Resources

Sample truss(1) Output

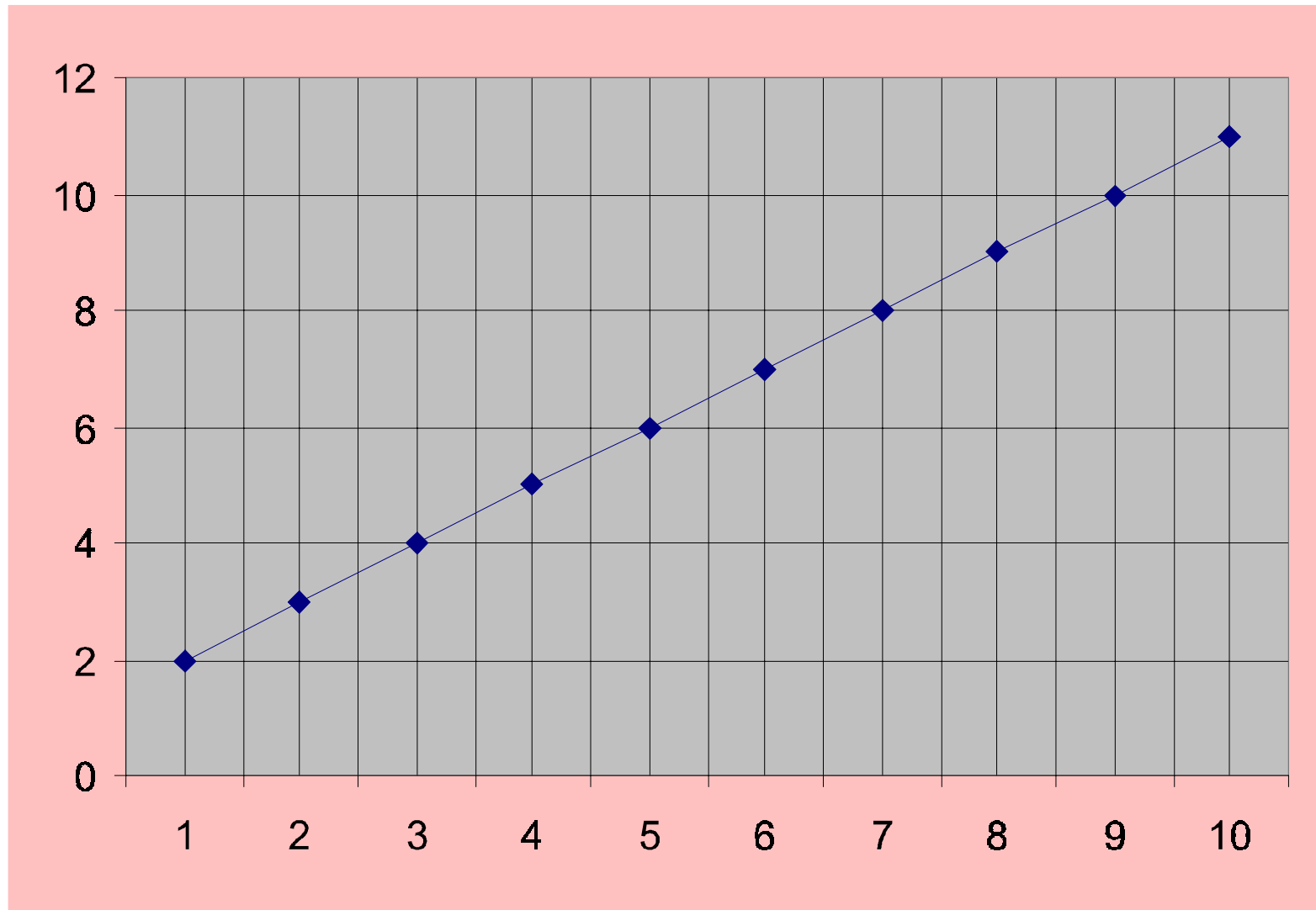


```
$ truss cat foo
execve("/usr/bin/cat", 0x08047BA0, 0x08047BAC)   argc = 2
brk(0x08304FBC)                               = 0
fstat(4, 1, 0x08047A60)                       = 0
open("foo", 02000000, 0666)                   Err#2  ENOENT
open("/usr/lib/locale/C/LC_MESSAGES/uxsyserr", 0_RDONLY, 01002767707) Err#2  ENOENT
open("/usr/lib/locale/C/LC_MESSAGES/uxlibc", 0_RDONLY, 01002767700) Err#2  ENOENT
UX:cat: write(2, " U X : c a t : ", 8)         = 8
ERRORwrite(2, " E R R O R", 5)                = 5
: write(2, " : ", 2)                          = 2
open("/usr/lib/locale/C/LC_MESSAGES/uxcore.abi", 0_RDONLY, 01001074600) Err#2  ENOENT
Cannot open foo: No such file or directory
write(2, " C a n n o t   o p e n   ".., 43)    = 43
_exit(2)
$
```

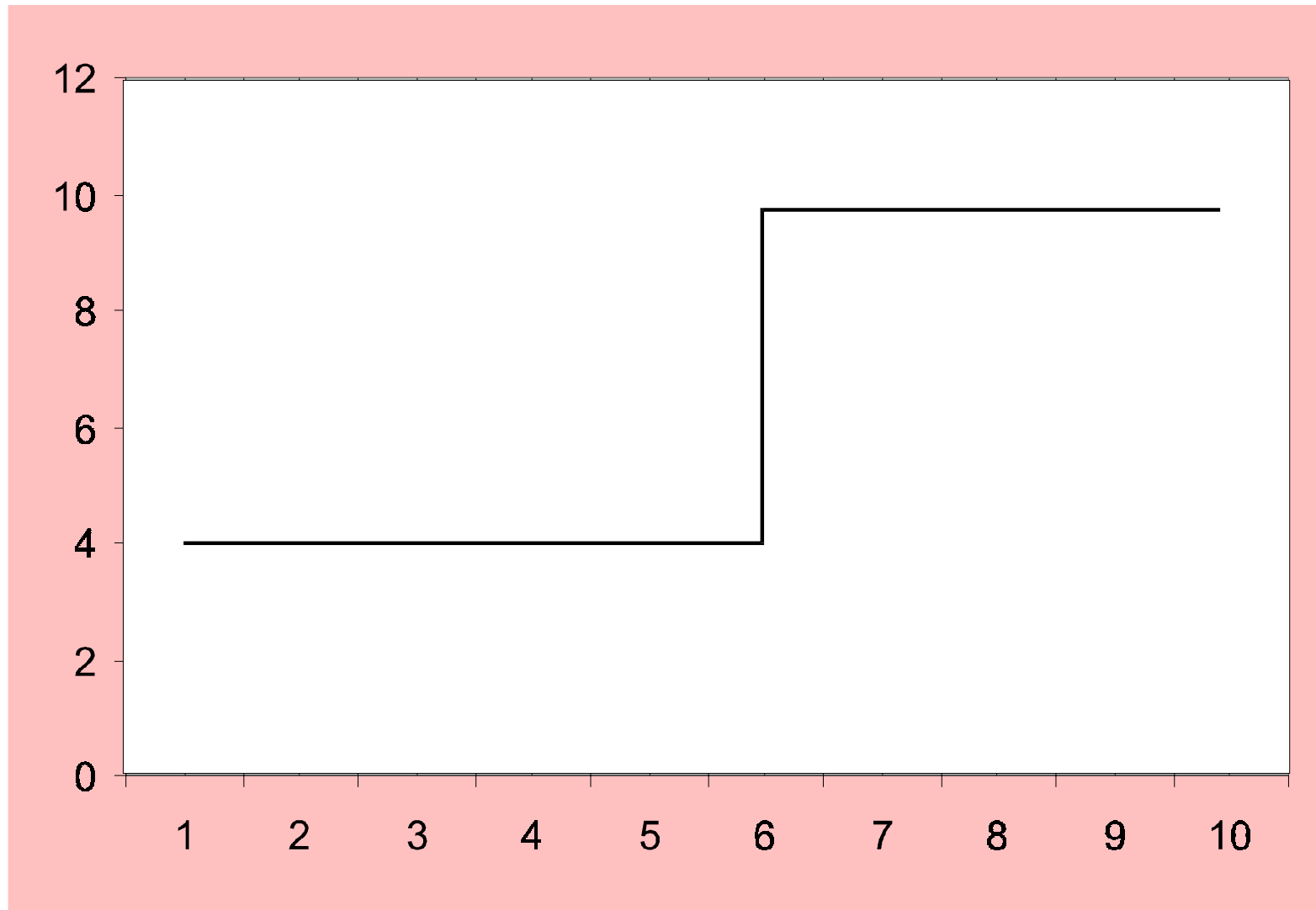
Tuning an NSC Kernel

- Types of Kernel Tunables
- Autotune
- Performance Monitoring Tools
- NSC-Specific Kernel Tunables
- Building Kernels and **idtune**(1M) in NSC

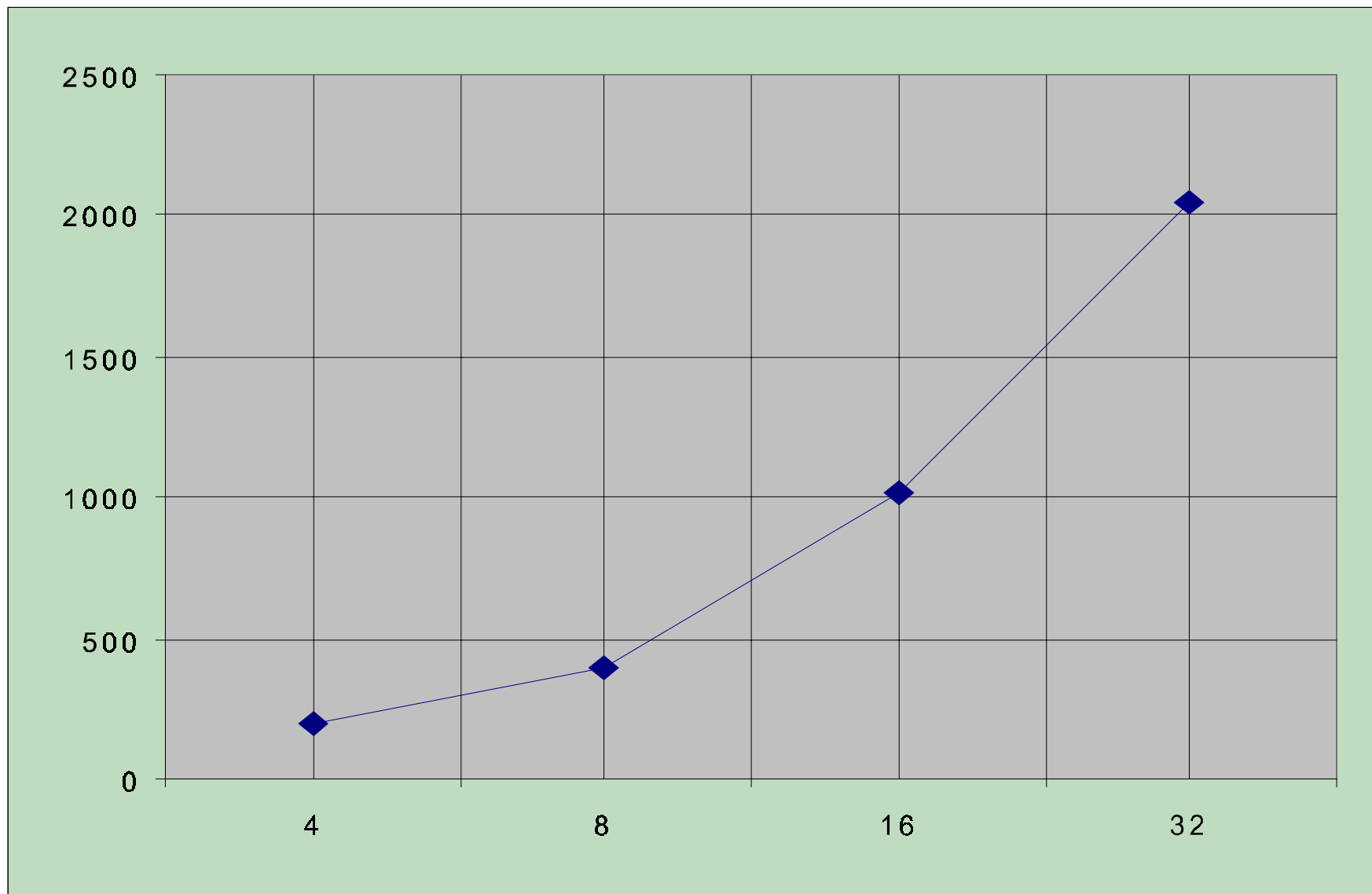
Graph of $Y=X+1$



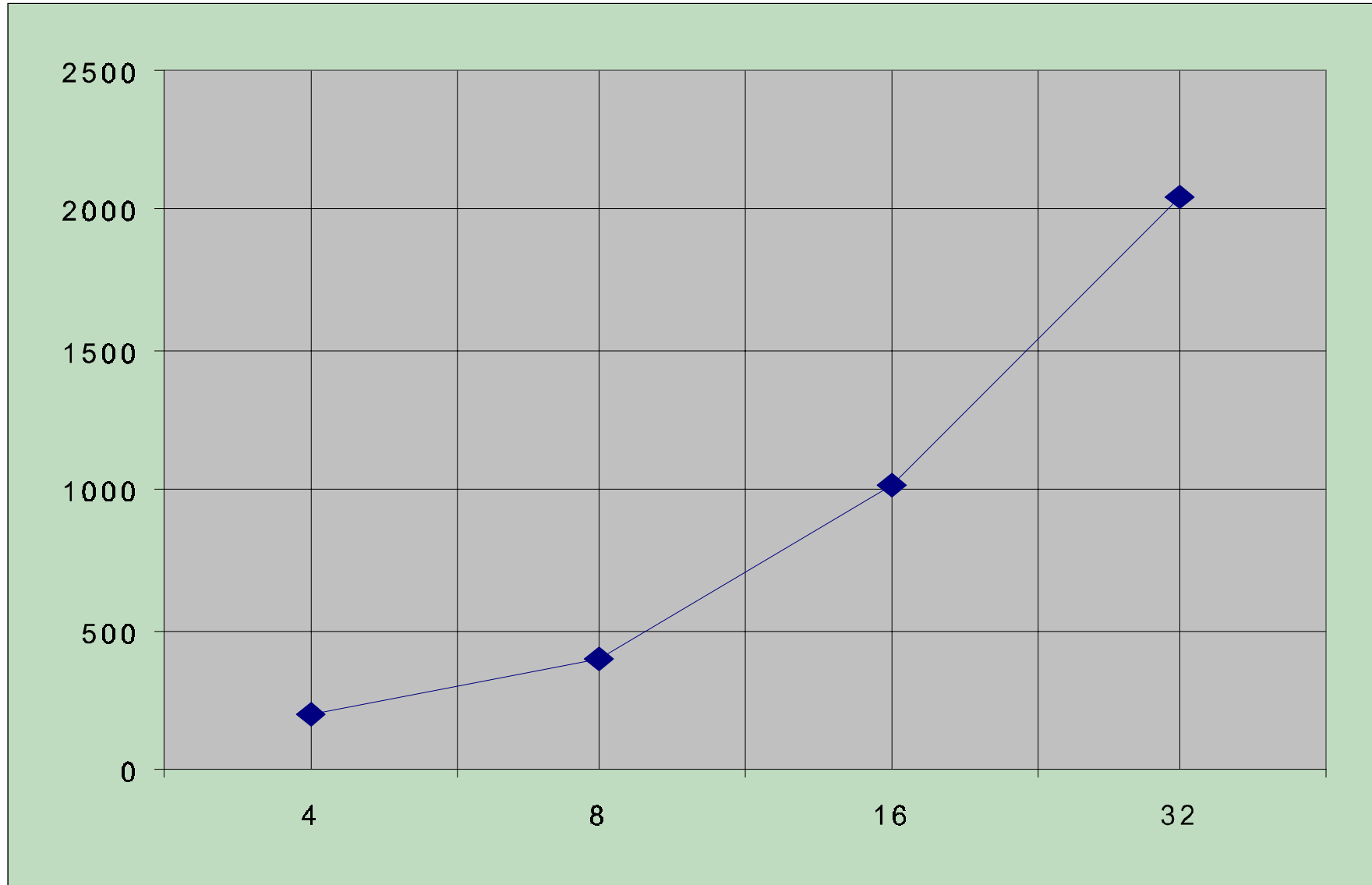
Graph of a Step Function



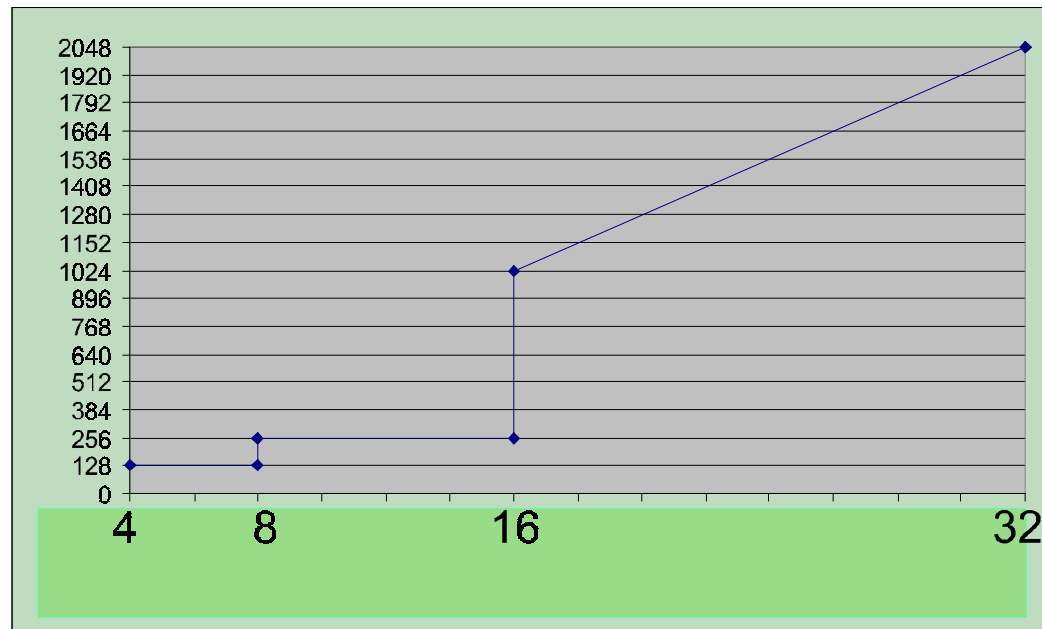
Example Curve of an Auto-Tuned Parameter



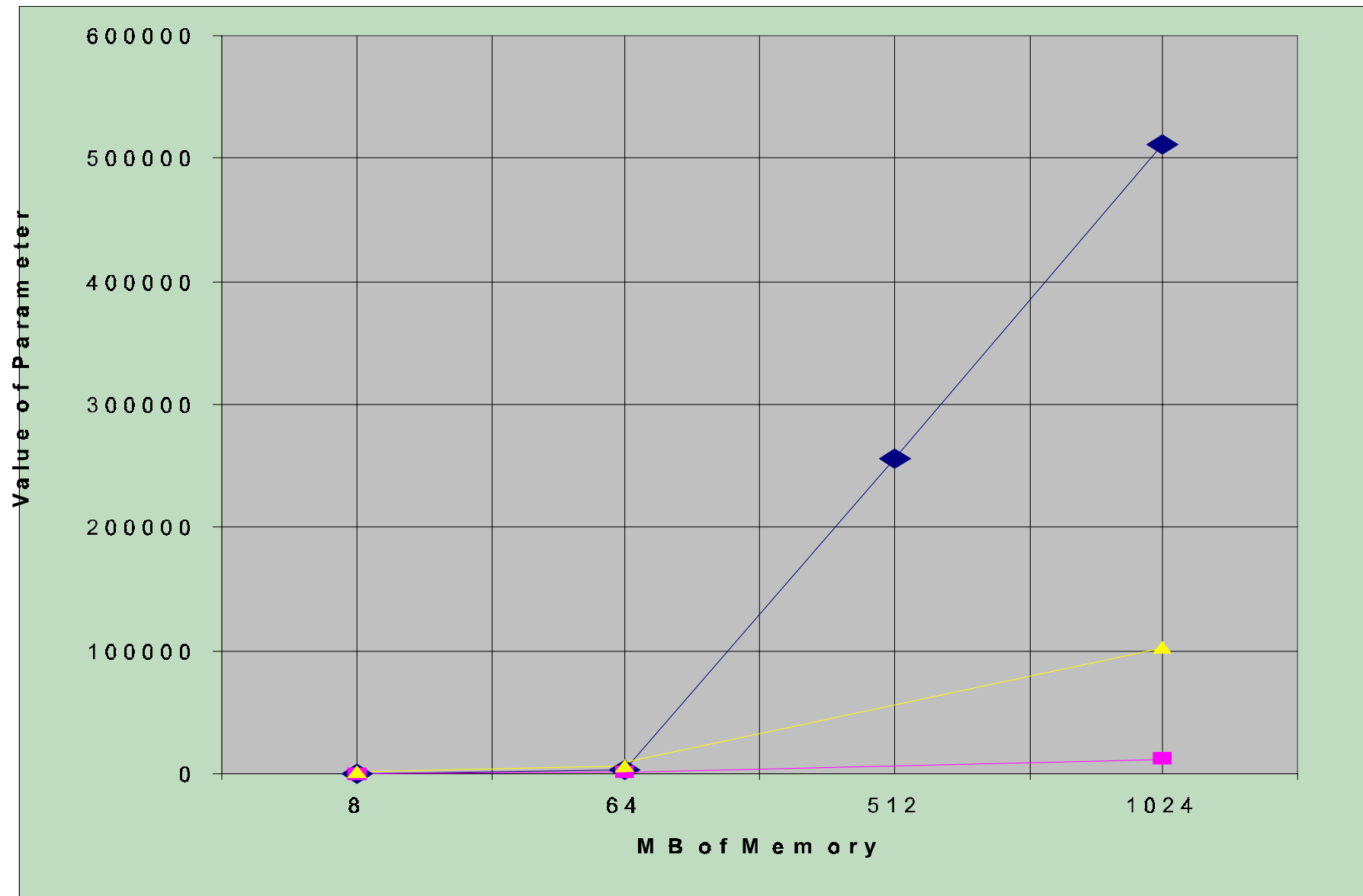
Example Curve of an Auto-Tuned Parameter



An Example Curve With Steps



A Real Example



NSC-Specific Kernel Tunables

The screenshot shows a 'System Tuner' window with a tab labeled 'NSC Parameters'. The window contains a list of kernel tunables on the left and a description on the right. Below the list is a progress bar and a set of control buttons.

Tunable	Value
NCNODE	300
CFS_ROOTBKUP_USERMEM	500
CFS_ROOTBKUP_KERNMEM	25
CLMS_NOEDOWN_FP	1
CLMS_FORCE_ROOT_DLY	300
CLMS_NDTO_WARN_SECS	150
CLMS_NDTO_PANIC_SECS	300
INET_ASYNC_DAEMONS	10
INET_PSVR_DEFAULT	1
INET_TCP_LHASHSZ	40

These tunable parameters are used for configuring the NSC kernel.

300

OK Reset Reset to Factory Cancel Help