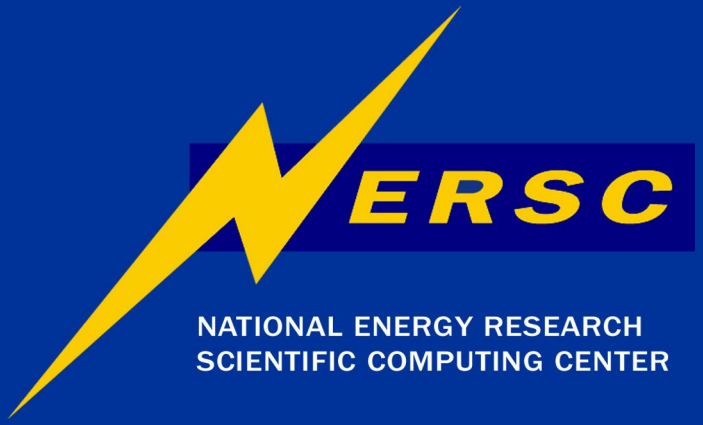

National Energy Research Scientific Computing Center (NERSC)

CHOS - CHROOT OS

Cary Whitney
NERSC Center Division, LBNL
September 27, 2004

Background

- PDSF is a medium size cluster used by a diverse group of High Energy and Nuclear Physics Groups
  - ATLAS
  - CDF
  - STAR
  - KamLAND
  - SNO
  - SNFactory (Astrophysics)

Motivation

Problem

Groups were starting to request different versions of RedHat (RH 7.2, RH 7.3, RH8)

Solution

CHOS - In house developed framework for supporting multiple Oss concurrently on a single system.

Requirements

- Support multiple OSs concurrently on each node
- Not require partitioning the cluster
- Be nearly transparent to the users
- Integrate with the batch/scheduler system
- Easily deployable across the cluster
- Scale with the number of requested OS releases

CHOS - CHROOT OS

- At its core, CHOS is chroot'ing into an alternate OS
- However, this alone isn't enough
  - File systems (both real and virtual)
  - Batch integration
  - Transparent and automatic
  - Scaleable for many OSs

Kernel Module

- Creates to files in proc file system (/proc/chos)
  - /proc/chos/link - Special symbolic link
  - /proc/chos/setlink - Writable file to set path for link
- /proc/chos/link has the following traits
  - Settable by setlink
  - Each process sees link pointing to its set value
  - Child processes inherit value of parent
- Following checks
  - Only root can set valid paths

CHOS link

```
[root@pc2622 root]# ls -l /proc/chos/
total 0
lrwxrwxrwx    1 root     root            1 Sep 24 20:40 link -> /
-rw-rw-rw-    1 root     root            0 Sep 24 20:40 resetchos
[root@pc2622 root]# cat /proc/chos/valid
/auto/common/os/redhat62
/auto/common/os/redhat8
/auto/common/os/redhat9
/auto/redhat73
/local/root
[root@pc2622 root]# echo /etc/redhat-lsb/
/etc/redhat-lsb/
[root@pc2622 root]# echo '/auto/redhat73' > /proc/chos/setchos
[root@pc2622 root]# cat /etc/redhat-release
Red Hat Linux release 7.3 (Valhalla)
[root@pc2622 root]# ls -l /proc/chos/
total 0
lrwxrwxrwx    1 root     root            1 Sep 24 20:41 link -> /auto/redhat73
-rw-rw-rw-    1 root     root            0 Sep 24 20:41 resetchos
```

CHOS Top Directory

```
[root@pc2622 root]# ls -l /chos
total 8
drwxrwxrwx    2 root     root         4096 Sep 22 14:57 afs
drwxr-xr-x    2 root     root            0 Sep 23 15:53 auto
lrwxrwxrwx    1 root     root           20 Sep 22 17:45 bin -> /proc/chos/link//bin
lrwxrwxrwx    1 root     root           20 Sep 22 17:45 dev -> /proc/chos/link//dev
lrwxrwxrwx    1 root     root           20 Sep 22 17:45 etc -> /proc/chos/link//etc
drwxr-xr-x    2 root     root         4096 Sep 22 17:45 export
lrwxrwxrwx    1 root     root           22 Sep 22 17:45 extra -> /proc/chos/link//extra
lrwxrwxrwx    1 root     root           21 Sep 22 17:45 home -> /proc/chos/link//home
lrwxrwxrwx    1 root     root           23 Sep 22 17:45 initrd -> /proc/chos/link//initrd
lrwxrwxrwx    1 root     root           20 Sep 22 17:45 lib -> /proc/chos/link//lib
drwxr-xr-x    2 root     root            0 Sep 23 15:55 local
lrwxrwxrwx    1 root     root           20 Sep 22 17:45 opt -> /proc/chos/link//opt
lrwxrwxrwx    1 root     root           10 Sep 22 17:45 proc -> local/proc
lrwxrwxrwx    1 root     root           21 Sep 22 17:45 sbin -> /proc/chos/link//sbin
lrwxrwxrwx    1 root     root           14 Sep 22 17:45 scratch -> /local/scratch
lrwxrwxrwx    1 root     root           15 Sep 22 17:45 tmp -> /local/root/tmp
lrwxrwxrwx    1 root     root           18 Sep 22 17:45 u -> /proc/chos/link//u
lrwxrwxrwx    1 root     root           20 Sep 22 17:45 usr -> /proc/chos/link//usr
lrwxrwxrwx    1 root     root           20 Sep 22 17:45 var -> /proc/chos/link//var
```

PAM Module

- PAM module that provide a “session” component
- PAM module looks at contents of .chos file in the user’s home directory
- Performs the necessary steps to initiate a CHOS session
- Can be added to PAM configuration for ssh to automatically use the alternate OS upon login

Batch Integration

- Modified job starters are used for that batch system
- Job starter looks for CHOS environmental variable
- Automatically switches if CHOS variable is set to a valid OS
- PAM module sets CHOS variable, so no further action is required by the user wanting to run the same OS

Use Cases

- Independently upgrading base OS without forcing users to switch platforms
- Provide test bed for users evaluating or migrating to new Oss.
- Provide access to older releases (un-maintained) in more secure fashion for re-running old codes or applications
- Run binaries compiled for a specific release in CHOS, while running other services in base OS

Security

- CHROOT is a privileged operation for a reason
- CHOS allows administrator to specify which alternate Oss are allowed
- CHOS checks against this list before initiating a CHOS session
- Services would typically be run out of just the base OS
- Disable setuid programs in alternate Oss to limit security risks. If application needs to be setuid, symlink to local installation

Current Status

- Tested with both 2.4 and 2.6 kernels
- Base OS: RedHat, SuSE, Fedora, Scientific Linux
- Alternate OS: RedHat, Fedora, Scientific Linux
- Tested with multiple versions of RedHat and SuSE

Future Work

- Simplified installation - Already in RPM format. Future release may automatically mount local file systems under CHOS
- PAM enabled job starter - Re-use PAM module for batch system as well. This job starter could have other uses (pam_limits).

Conclusion

- Dealing with competing requirements from users is a typical problem for shared resources
- CHOS greatly diminishes this problem for providing various operating systems
- CHOS also helps decouple the needs of the system administrator from the needs of the user

Availability

- Software available at:
  - http://www.nersc.gov/nusers/resources/PDSF/chos/
- Additional information:
  - Email: canon@nersc.gov