The CLuED0 Linux Cluster

CHEP 2003

Bill Lee, Roger Moore and Dugan O'Neil

Florida State University and Michigan State University

Outline

- Introduction
- Cluster Configuration
- Cluster Management Software (CLuMP)
- Administrative Model
- Code Development Platform
- Batch Processing Farm (PBS)
- Data Access (SAM)
- Summary

Introduction

- CLuED0 is the cluster of all DØ Linux desktop machines.
- Currently 250 nodes (and growing) from 50 institutes with over 500 users.
- Unique management tools and management model have been developed. Management by users.
- Primarily a desktop cluster, but also provides significant functionality for data access, code development, batch processing, etc.

History

- DØ Computing Early 2000
  - d0mino - a 176-processor SGI Origin 2000 system
  - SGI cluster
  - Windows NT cluster with main central servers
  - DØ Fermi Linux cluster
  - Unclustered self-managed Linux PCs
  - Various X terminals

CLuED0 Beginnings

- Michigan State University (MSU) Post Docs decide to cluster some MSU Linux desktops together.
- Other groups join, first in the same building, then spreading to other buildings. CLuED0 is born.
  - The initial goal of CLuED0 is to provide the advantages of a cluster while still providing users the ability to configure their desktop.
- Fermilab requires a single Linux cluster at DØ for security reasons.
- All Linux PCs at DØ are now in the CLuED0 cluster.

Cluster Configuration

- Currently RedHat 7.1 based cluster, planned upgrade to RedHat 8.1 when available.
- Machines in 6 different buildings on Fermilab site.
- One rack of servers in central location at DØ provides web, LDAP, batch and data access services. Location for institute-owned disk servers.
- Home directories and DØ code distributions mounted from DØ central services (8 processor SGI). DØ provides nightly backup.
- Slave LDAP servers in each building with a failover chain to all other buildings.

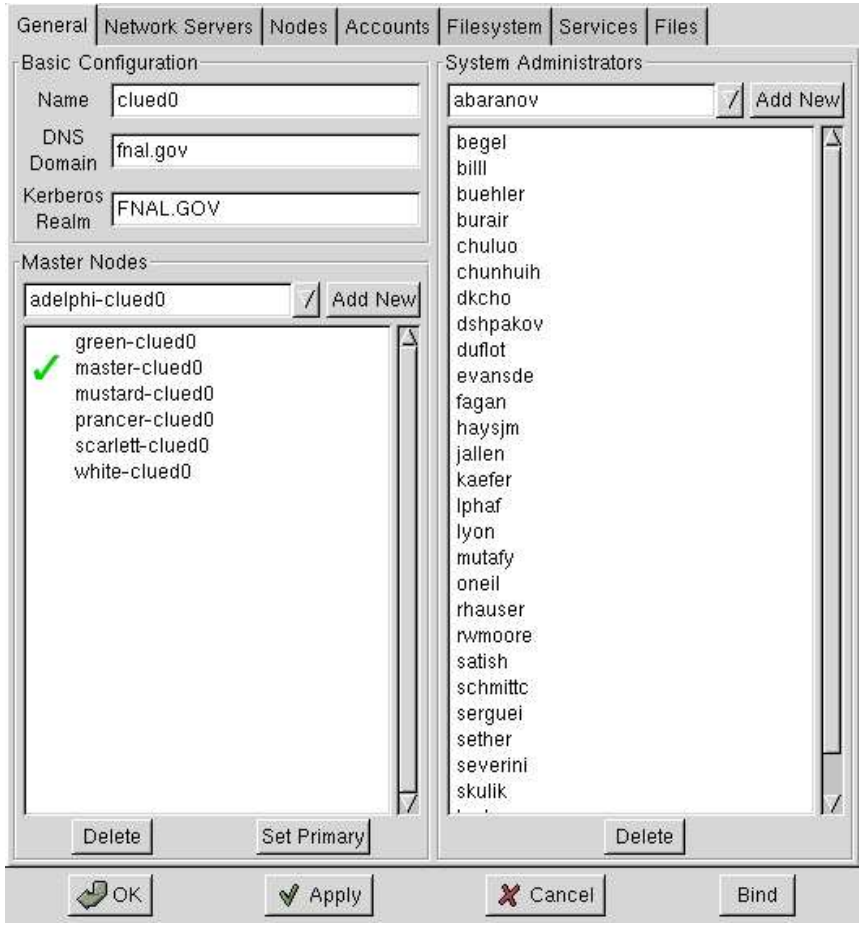
CLuMP

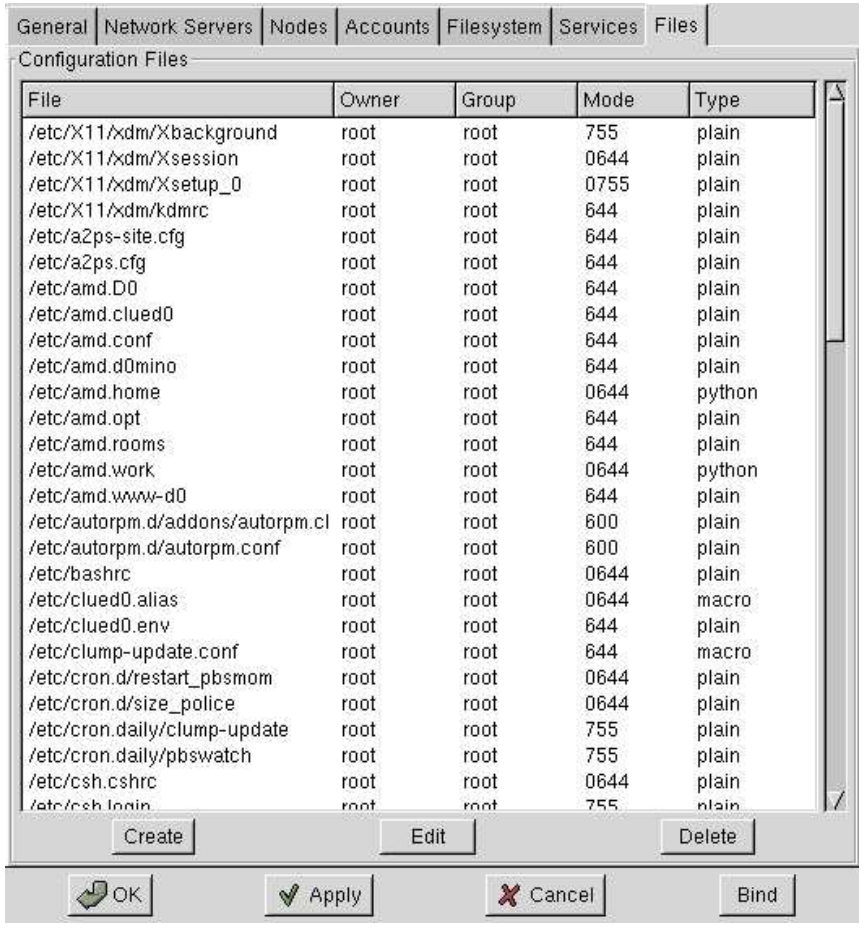
- CLuED0 is NOT a homogeneous system
- Very diverse hardware
  - P2,P3,P4,AMD,Cyrix
  - Speed 200MHz-2.4GHz
  - Memory 64Mb-4Gb
  - Disk 6Gb-2.5Tb
- Diverse usages and priority functionality (5O institutes).
- Configurations in central database (currently LDAP, soon moving to mysql). Custom schema and management software (CLuMP).

CLuMP

- CLuMP allows us to tell LDAP about the structure of our cluster
  - Configuration for cluster, netgroup, nodes.
  - Store configuration files at all three levels (local overrides general)
  - built-in users, autofs config
- Configuration files can also be automatically generated from the database using python scripts. For example, /etc/hosts dynamically generated from global list of nodes.
- Provides command line and GUI interfaces.

CLuMP


General
Network Servers
Nodes
Accounts
Filesystem
Services
Files
Basic Configuration
Name
clued0
DNS Domain
fnal.gov
Kerberos Realm
FNAL.GOV
Master Nodes
adelphi-clued0
Add New
green-clued0
master-clued0
mustard-clued0
prancer-clued0
scarlett-clued0
white-clued0
Delete
Set Primary
System Administrators
abaranov
Add New
begel
billl
buehler
burair
chuluo
chunhuih
dkcho
dshpakov
duflot
evansde
fagan
haysjm
jallen
kaefer
lphaf
lyon
mutafy
oneil
rhauser
rwmoore
satish
schmittc
serguei
sether
severini
skulik
Delete
OK
Apply
Cancel
Bind


General
Network Servers
Nodes
Accounts
Filesystem
Services
Files
Configuration Files
File
Owner
Group
Mode
Type
/etc/X11/xdm/Xbackground root root 755 plain
/etc/X11/xdm/Xsession root root 0644 plain
/etc/X11/xdm/Xsetup_0 root root 0755 plain
/etc/X11/xdm/kdmrc root root 644 plain
/etc/a2ps-site.cfg root root 644 plain
/etc/a2ps.cfg root root 644 plain
/etc/amd.D0 root root 644 plain
/etc/amd.clued0 root root 644 plain
/etc/amd.conf root root 644 plain
/etc/amd.d0mino root root 644 plain
/etc/amd.home root root 0644 python
/etc/amd.opt root root 644 plain
/etc/amd.rooms root root 644 plain
/etc/amd.work root root 0644 python
/etc/amd.www-d0 root root 644 plain
/etc/autorpm.d/addons/autorpm.cl root root 600 plain
/etc/autorpm.d/autorpm.conf root root 600 plain
/etc/bashrc root root 0644 plain
/etc/clued0.alias root root 0644 macro
/etc/clued0.env root root 644 plain
/etc/clump-update.conf root root 644 macro
/etc/cron.d/restart_pbsmom root root 0644 plain
/etc/cron.d/size_police root root 0644 plain
/etc/cron.daily/clump-update root root 755 plain
/etc/cron.daily/pbswatch root root 755 plain
/etc/csh.cshrc root root 0644 plain
Create
Edit
Delete
OK
Apply
Cancel
Bind

Administrative Model

- No computing professionals available “on the ground”. Computing division maintains home directories, backups, central code repository, but do not touch clustered Linux machines.
- CLuED0 administration is provided solely from volunteer administrators (users) committed to contribute approximately O.2FTE. Recognized as official DØ service work.
- Any user can become an administrator simply by volunteering to take a responsibility.

Administrative Model

- Philosophy is to have at least 2 people identified in each building who know what they are doing and have root access to all machines. Users should be very comfortable approaching their local admins.
- Also designate a local contact person for each institute who is permitted to have root access to all the nodes belonging to that institute.
- Institutes buy/manage their own local disks (scratch areas) and are responsible for backing up their disks (or not). Very cheap place to put disk.

Code Development Platform

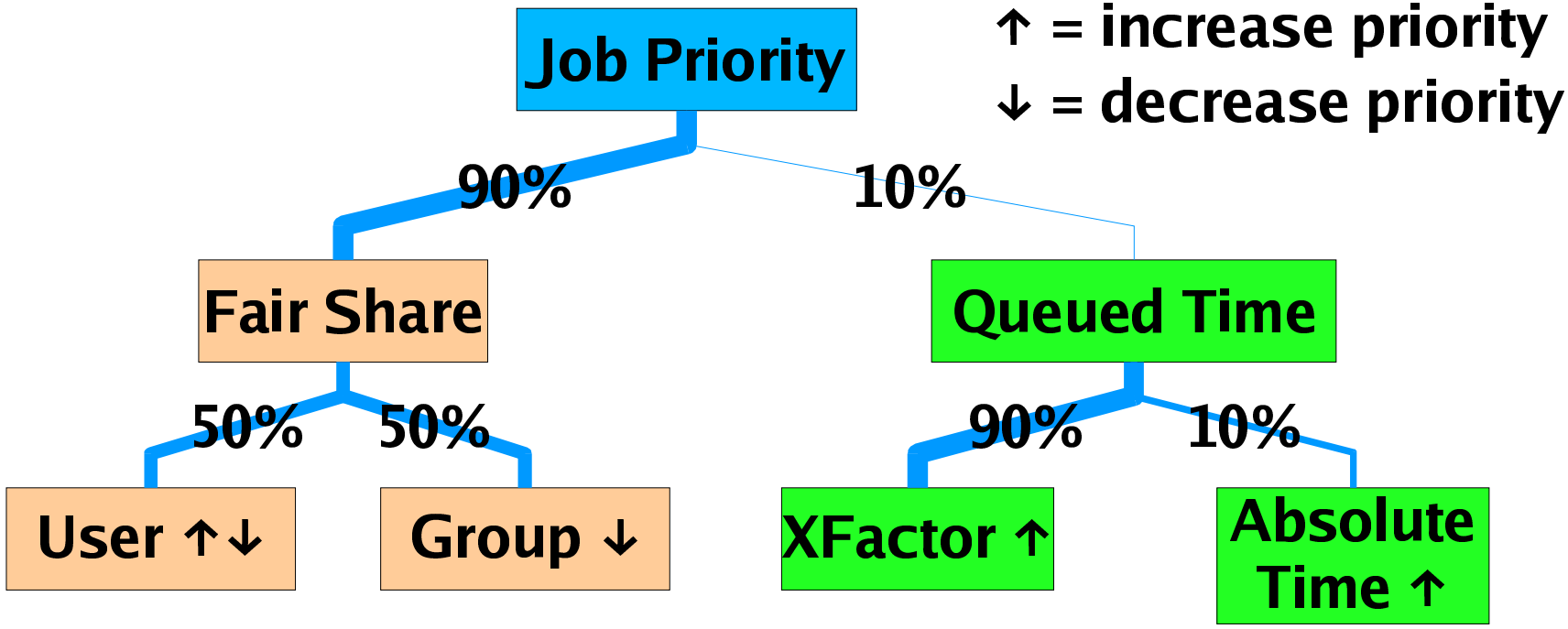
- Single most important code development platform for the experiment.
- Nightly builds of DØ code distributions are available on every machine via an NFS mount of the centrally maintained build-disk.
- Executables compiled on CLuED0 machines can be seamlessly sent to Linux-based central computing facilities.
- Besides basic desktop interface this is the most important cluster functionality.

Batch Processing Farm

- Run batch system on all nodes. OpenPBS with MAUI scheduler.
- So far has provided the majority of the analysis CPU used by the experiment (central analysis facility now also available)

Fairshare

- Fairshare employed by institute and by user. Most of batch priority is assigned based on usage from institute relative to institute contribution to the cluster (CPU power).



Data Access (SAM)

- All DØ data access is managed through SAM (Sequential Access via Metadata). See other talks at this conference.
- CLuED0 is a SAM station currently capable of transferring approximately 1Tb per day from central services to the cluster. Can be upgraded.
- Central 1Tb disk cache in central rack at the DØ
- Data transferred from tape or other stations to CLuED0 central cache. Client nodes then transfer (rcp) files to SAM-managed cache on local nodes. Interfaced to PBS.

Summary

- CLuED0 is a large desktop cluster at DØ (Fermilab).
- Primary functionality is desktop (email, web, office) but has been invaluable as code development and batch processing resource.
- Custom management software written (CLuMP). VERY useful for centralizing cluster configuration.
- Administration done by users (physicists) not computing professionals.