M08

Systems Management Using Linux

Terry Flickinger – ATS Linux Team

IBM @server xSeries Technical Conference

Aug. 9 - 13, 2004

Chicago, IL

Agenda

- Differentiate between Linux Cluster Management and Non-Cluster Management
- The three main components of Cluster management
  - Installation
  - Management
  - Monitoring

Agenda

- Tools for Cluster Management
  - xCat
  - CSM
- Tools for Non-Cluster Mangement
  - IBM Director 4.2
  - RDM

Agenda

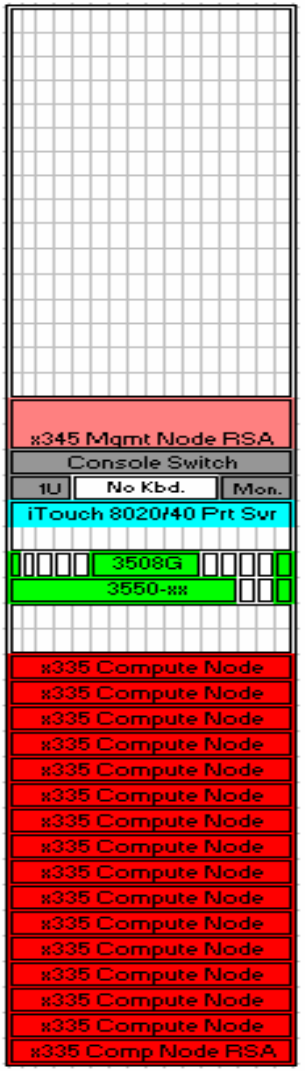
- Other Tools for Systems Management
  - ASU/LFLASH
- Which Tools to Recommend
- Resources

Cluster Mgmt v. Systems Mgmt

- Systems Management
  - One to One
  - Single operation applied to a single resource
- Cluster Management
  - One to Many
  - Single operation applied to multiple resources in parallel
- In the linux cluster space we are managing dozens, hundreds or thousands of servers
- GUI point and click to single nodes takes too much time, not practical
- Tools must be scriptable, command line, and operate across ranges of nodes
  - ex. “rpower rack01 off”
  - ex. “rinstall node01-node44”

Typical Linux Cluster Layout

- Mgmt/Head Node
  - 2 NICS… one to private cluster network. 2nd to public network
  - Serves DHCP to private only
- Cluster Nodes
  - Private network connection
  - Only accessible through Mgmt/Head node



Installation of Nodes

- If you look at all the Linux installation packages, you see the common denominator of loading the machines from the network
- Most packages handle this by using PXELinux

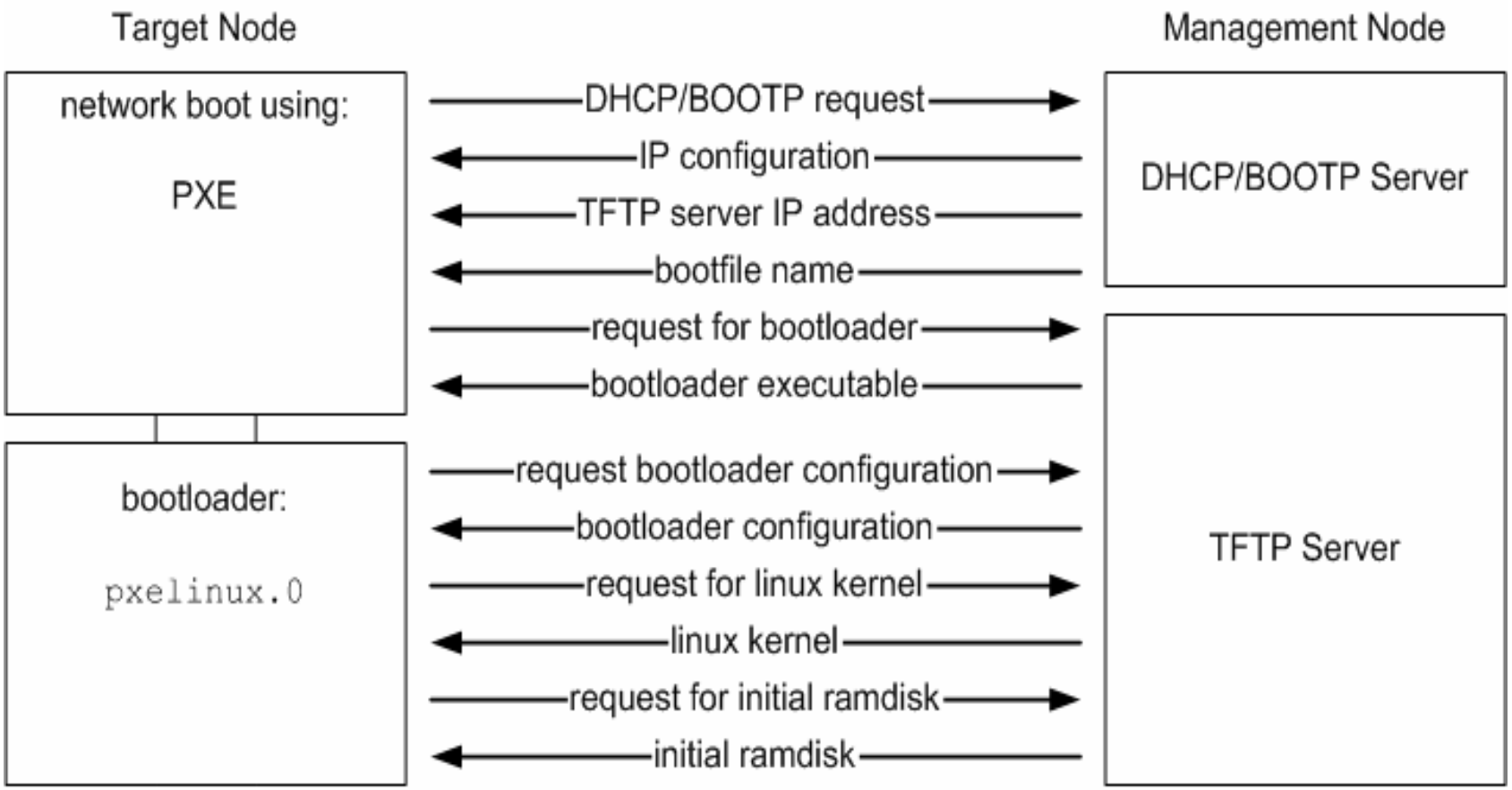
What is PXE ?

- PXE = “Preboot Execution Environment”
- Industry standard interface that allows a computer to boot off the ROM of the network interface
- Client/Server – The client gets a DHCP address and the Server (DHCP server) sends an address and a list of machines to boot from
- Client then gets boot image via TFTP

PXELINUX

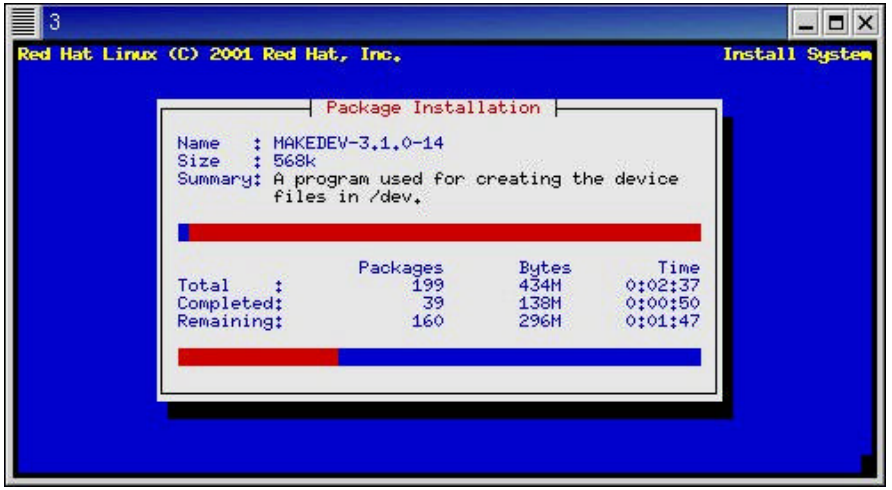
- PXELINUX is a SYSLINUX derivative for booting linux off a network server
- Every Linux install package uses this as the “core” of the installation procedure
- A basic understanding of this process will help you understand all the methods discussed
- http://syslinux.zytor.com/

PXE Installation Process



Kickstart Install

- Kickstart – RedHat's automated unattended installation procedure
- Similar to Windows Unattended install
- Kickstart file defines all installation parameters and provides answers to installation questions



SystemImager

- Part of the System Installation Suite of tools
- Creates source of target machine, and then pushes out to multiple clients using rsync
- PXE process exactly the same as described above except loads SystemImager kernel
- Other tools in System Installer suite to update the image and nodes without re-imaging

Partition Imager

- Open Source package that is very similar to Ghost
- Select target machine to collect image from
- Push image out to clients
- Image is a tarball and can be manipulated if necessary to apply updates
- Machine must be reimaged to apply changes

Tools to Make Installation Easier

- Tools exist that script and wrap around the open source tools
  - Creation of /tftpboot/pxelinux.cfg files
  - Creation and maintenance of dhcpd.conf file
  - Collection of MAC addresses
  - Creation of DNS files

xCAT

- E**x**treme Cluster Administration Toolkit
- Swiss Army knife for cluster installation and management
- Collection of scripts to front-end all the open source tools

xCAT

- xCAT's architecture and feature set have two major drivers
  - **Real world requirements -** The features in xCAT are a result of the requirements met in hundreds of real cluster implementations. When users have had needs that xCAT or other cluster management solutions couldn't meet, xCAT has risen to the challenge. Over the last few years, this process has been repeatedly applied, resulting in a modular toolkit that represents best practices in cluster management and a flexibility that enables it to change rapidly in response to new requirements and work with many cluster topologies and architectures.
  - **Unmatched Linux clustering experience -** The people involved with xCAT's development have used xCAT to implement many of the world's largest Linux clusters and a huge variety of different cluster types. The challenges faced during this work has resulted in features that enable xCAT to power all types of Linux clusters from the very small to the largest ever built.

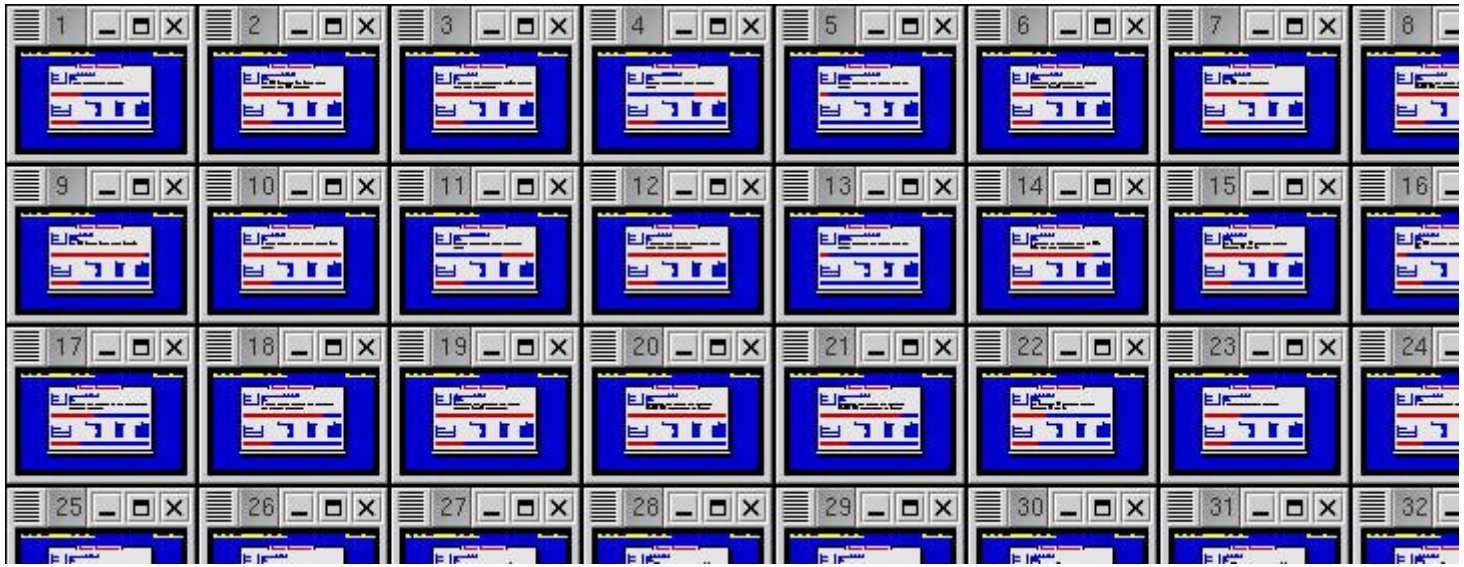
-

- Taken from www.xcat.org

xCAT Node Installation

- xCAT scripts manipulate the /tftpboot directory to handle the PXE installation
  - nodeset - changes the /tftpboot/pxelinux.cfg files for installation or different stages
  - mkks - generates the kickstart files, ramdisk images, and kernels for installation
  - rinstall, winstall - runs nodeset, opens window to watch install, and resets power on the nodes

xCAT Node Installation



- wcons windows allow monitoring of installation
- If a problem occurs, you can take a closer look

xCAT Tools

- Extensive use of the IBM Service Processor
  - rinv - Hardware information and VPD
  - rvitals - Hardware status
  - rpower - Power control
- Parallel commands
  - psh – parallel shell to run commands across nodes
  - prcp – remote copy
  - prsync – remote sync

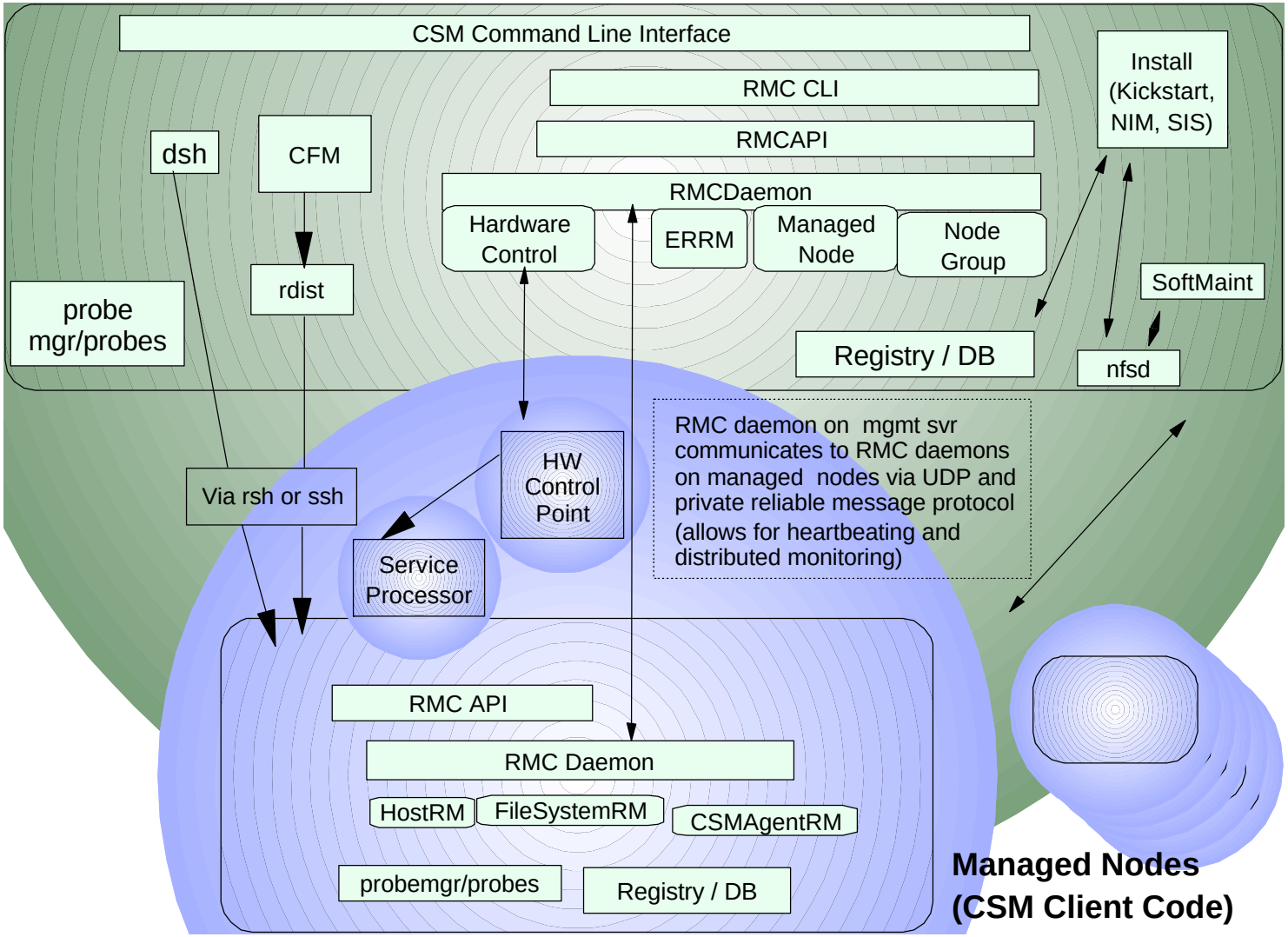
CSM

- Leverages IBM's supercomputing expertise and focus on Linux
- Concepts and technology derived from IBM Parallel System Support Programs for AIX (PSSP) and from open source tools
- Aimed at customers who want robust cluster systems management on an open, Intel-based server platform

CSM

- Manage multiple node from one system
  - Take control of remote nodes
  - Add, remove or change nodes
  - Run remote commands across nodes or groups of nodes
  - Synchronize configuration files across multiple nodes
- Automate operations
  - Monitor critical system components
  - Hardware and software status
  - Set events to run when selected criteria is reached

CSM Architecture



CSM

- Distributed Management Server
  - Monitor multiple nodes as if they were one
  - Manage groups of nodes
  - Maintain a persistent repository of information on each node
  - Maintain a heartbeat status of each node
- Remote Hardware Control
  - Take control of another node
  - Remotely power the node on, off or reset
  - Utilizes the IBM Systems Management Processor
- Distributed Shell
  - Run commands or scripts remotely
  - Combine the output from multiple servers

CSM

- Event Response Resource Manager
  - Run commands or scripts in response to user defined events
  - Includes a rich set of predefined conditions and responses
  - Monitor many system resources, including:
    - File systems
    - Programs
    - Adapters
    - Node availability
- Set alerts and responses to events
  - Notification
  - Application launch
  - Script launch
- Probe Manager
  - Check consistency of cluster configuration information
  - Diagnose errors

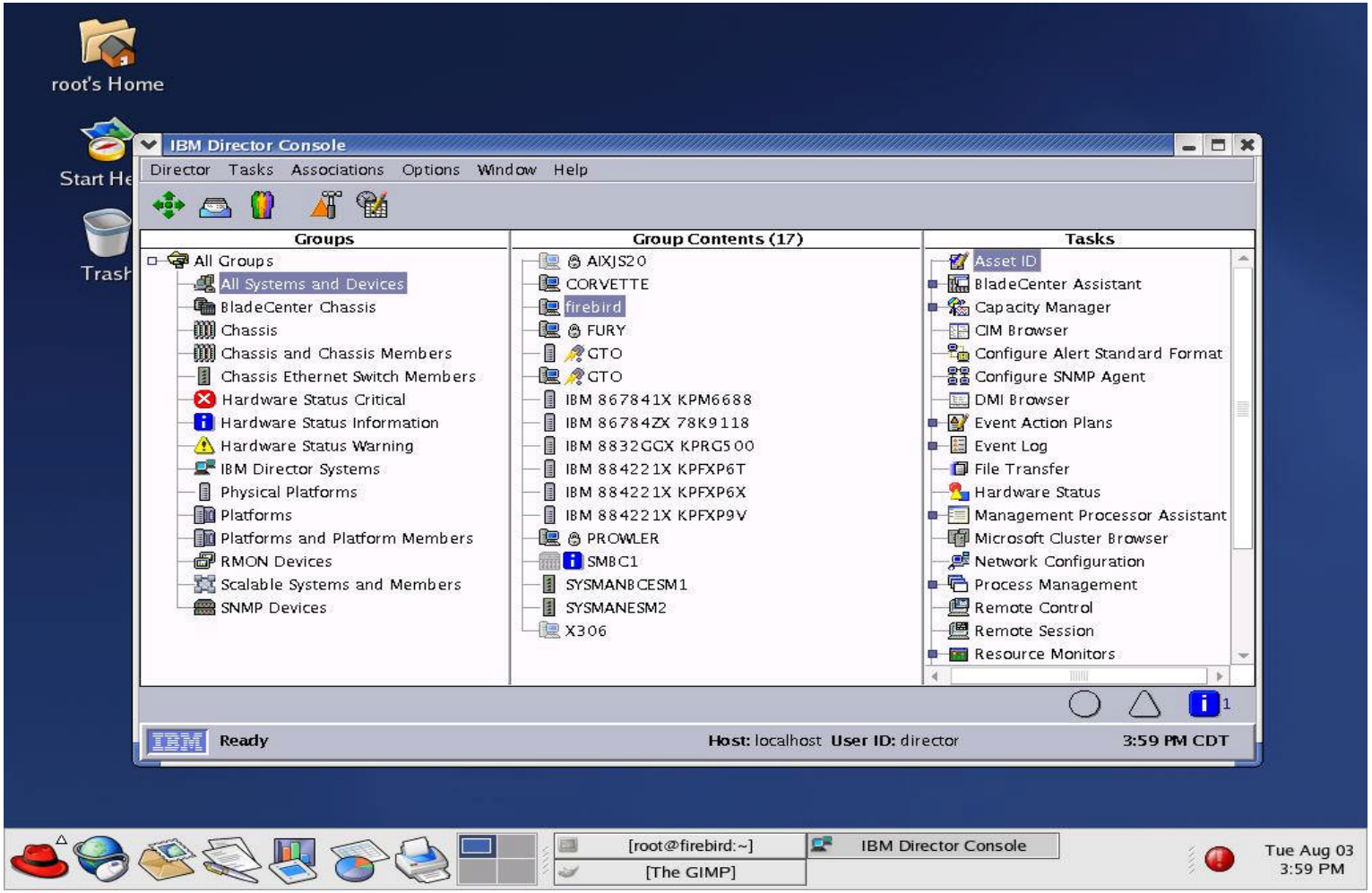
CSM

- Configuration File Management (CFM)
  - Repository for files that are common between the cluster nodes
  - Central location of files
    - Make changes to common files in one location
  - Changes are synchronized throughout the cluster
  - Each managed node can pull the changes at boot time, when files change, and/or when administrator runs a command on the managed node

CSM Pros/Cons

- Official supported IBM Product
  - Phone number to call
  - Escalation paths with development
- Mixture of compiled binaries and scripts
  - Not Open Source
- No tools to assist in HPC cluster setup
- Heterogeneous management of Linux and AIX clusters from one console

Non-Cluster Management
IBM Director 4.2

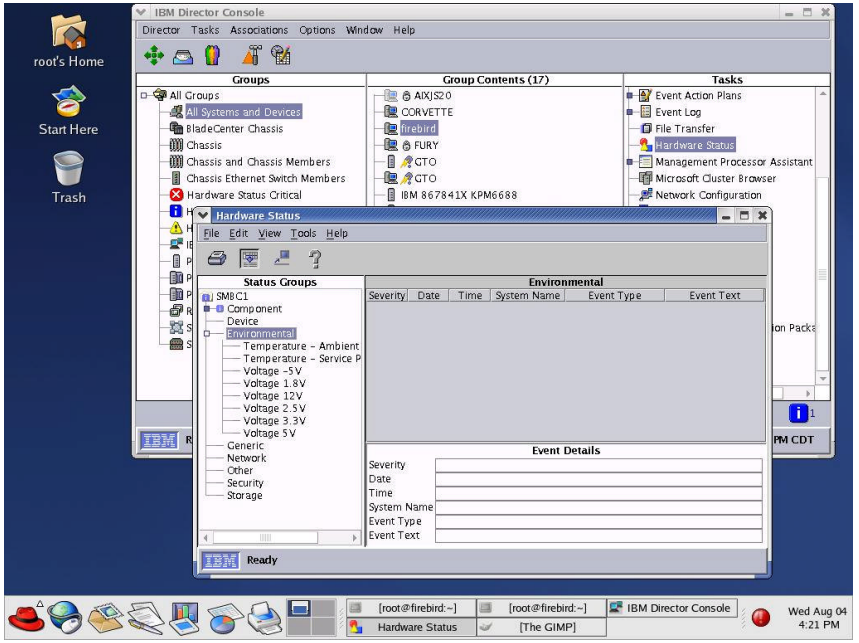


IBM Director 4.2

- Similar in functionality to Windows based software
- Allows control/management of Windows and Linux based units
- Very robust, enterprise ready

Remote Deployment Manager

- RDM can be used to install Linux to workstations and servers
- Uses kickstart to install OS.
- Future release will run on linux platforms



When to recommend IBM Director

- Like a Windows style of administration
- Want to use only a GUI for system management
- Want to manage Windows and Linux together
- Need the most complete PCI, ServeRAID, and BladeCenter hardware discovery and configuration support
- Want to support non-IBM hardware as well as IBM hardware
- Want higher level management functions like complete hardware/software Inventory, Capacity Manager (xSeries only), Application Workload Manager (xSeries only)

When to recommend CSM

- Want a command line and scripting ability for system administration
- Linux-only shop (Note that for IBM Director, Windows is required for deployment and pSeries Linux is not yet supported for deployment)
- Want to manage Linux and AIX together
- Want to manage pSeries and xSeries (Linux) together
- Need their systems management solution to fit into their specific environment (i.e., JPMC chose CSM over Director because of the way CSM could fit into their layered build environment)
- High Performance Computing shops, especially those running the Cluster 1350, would want CSM

Resources

- Websites
  - www.xcat.org
  - www.ibm.com/servers/eserver/clusters/software
  - ganglia.sourceforge.net
  - oscar.sourceforge.net
  - www.beowulf.org
  - www.top500.org