Cluster Management	Cluster Management L-Outline	
	Common Management Tools OSCAR ROCKS Other Popular Cluster	Regular Expression Meta-characters Regular Expression Meta-characters (cont.)
Cluster Management	Management tools Software Management/Change Control	SEC Logsurfer+ Security plans/procedures, Risk
James E. Prewett	Cfengine Getting Started with Cfengine Parallel Shell Tools / Basic Cluster	Analysis Network Topologies and Packet Filtering
October 8, 2008	Scripting PDSH Dancer's DSH Clusterit	Linux Tricks Cluster-specific issues Checking Your Work
	C3 tools (cexec) Basic Cluster Scripting Backup Management Logging/ Automated Log Analysis	Regression Testing System / Node / Software Change Management Logs How to know when to upgrade, trade–offs
< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □	Regular Expression Review	Monitoring tools @ ► < E ► < E ► E • OQC

Cluster Management	Cluster Management
Common Management Tools	Common Management Tools
Loscar	LOSCAR

OSCAR Information

OSCAR cluster distribution features:

Vital Statistics:		
Version:	5.1	
Date:	June 23, 2008	
Distribution Formats:	tar.gz	
URL:	http://oscar.openclustergroup.org/	

- Graphical Installation and Management tools
 - ... if you like that sort of thing

► Supports X86, X86_64 processors

Supports Ethernet networksSupports Infiniband networks

OSCAR (key) Cluster Packages

Whats in the box?

- ► Torque Resource Manager
- Maui Scheduler
- ► c3
- ► LAM/MPI
- MPICH
- OpenMPI
- OPIUM (OSCAR User Management software)
- pFilter (Packet filtering)
- PVM
- System Imager Suite (SIS)
- Switcher Environment Switcher

OSCAR Supported Linux Distributions

- RedHat Enterprise Linux 4
- RedHat Enterprise Linux 5
- Fedora Core 7
- Fedora Core 8
- ▶ Yellow Dog Linux 5.0
- OpenSUSE Linux 10.2 (x86_64 Only!)
- "Clones of supported distributions, especially open source rebuilds" of Red Hat Enterprise Linux such as CentOS and Scientific Linux, should work but are not officially tested."

Cluster Management Cluster Management Common Management Tools Common Management Tools

OSCAR Installation

- Install a supported Linux on the erver Node Leave at least 4GB free in each of / and /var! The easy way is to make 1 big partition for / !
- Create repositories for SystemInstaller # mkdir /tftpboot # mkdir /tftpboot/oscar # mkdir /tftpboot/distro # mkdir /tftpboot/distro/OS-version-arch
- Unpack the oscar-repo-common-rpms and the oscar-repo-DISTRO-VER-ARCH tarballs into /tftpboot/oscar/
- Copy your RPMs into the /tftpboot/distro/OS-version-arch directory

OSCAR Installation (cont.)

- Install yum unless your OS already has it
- Install yume: # yum install createrepo /tftpboot/oscar/common-rpms/yume*.rpm
- Install oscar-base RPM: # yume -- nogpgcheck¹ -- repo /tftpboot/oscar/common-rpms install oscar-base

▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● の Q @

◆□▶ ◆□▶ ◆三▶ ◆三▶ ● □ ● ○○○

¹This is not in the documentation, but I found that the packages were not signed causing yume to barf unless you passed it the --nogpgcheck option. YMMV

OSCAR Server Node Network Configuration

- Give your host a hostname! The default of "localhost" or "localhost.localdomain" will *not* work.
- Configure the "Public" network interface as per the requirements of your local network. This is the network that will connect to the Internet (or the lab network), so configure it appropriately.
- Configure the "Private" network interface using a "Private" IP address.

The IANA has reserved the following three blocks for private internets:

- 10.0.0.0 10.255.255.255 (10/8 CIDR block)
- 172.16.0.0 172.31.255.255 (172.16/12 CIDR block)
- 192.168.0.0 192.168.255.255 (192.168/16 CIDR block)

Step 8:

Cluster Management Common Management Tools OSCAR

OSCAR Cluster Installation

Once the Server is installed and configured, start the installer! # cd /opt/oscar # ./install_cluster <device> This will:

- ► Install all required RPMs
- update the /etc/hosts file with OSCAR aliases
- update the /etc/exports file
- update system initialization scripts (/etc/rc.d/init.d/)
- restart any affected services

Then the installer GUI will be launched.

Reset

Build Image

▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● の Q @ ▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● の Q @ **Cluster Management Cluster Management** Common Management Tools Common Management Tools The OSCAR Installation Wizard: Build Client Image X OSCAR Wizard - headnode 000 Choose an image name Specify Disk Partition file Chose a package file Welcome to the OSCAR Wizard Pick IP assignment method OSCAR Version: 5.1 Chose a Target Distribution - INSTALL MODE -Specify package repositories Pick Post Install action Select your packages Help Step 0: Sten 1: Select OSCAR Packages To Install... Help Build OSCAR Client Image - I = X Configure the packages Configure Selected OSCAR Packages.. Step 2: Help Fill out the following fields to build a System Installation Suite image. If you need help on any field, click the help Step 3: Install OSCAR Server Packages Help ► Install the Server packages button next to it Build OSCAR Client Image.. Step 4: Help Image Name: oscarimage Help Step 5: Define OSCAR Clients. Help Build an image for the Package File: /opt/oscar/oscarsamples/rhel Choose a File.. Help Step 6: Setup Networking... Help compute nodes Target Distribution: redhat-el-as-4-i386 Help Delete OSCAR Clients... Help Monitor Cluster Deployment Help Package Repositories: Help /tftpboot/oscar/common-rpms Define the compute nodes Before continuing, network boot all of your nodes. Once they have completed installation, reboot them from /opt/oscar/oscarsamples/ide. Choose a File.. **Disk Partition File:** Help the hard drive. Once all the machines and their ethernet Configure networking daptors are up, move on to the next step **IP Assignment Method:** static Help **Complete Cluster Setup** Step 7: Help Post Install Action: reboot Help

- Complete the setup
- Test the cluster!

Help

Test Cluster Setup

Quit

Close

Cluster Management Common Management Tools OSCAR

Define OSCAR Clients (Compute Nodes)

- Pick the image to install
- Specify the domain name
- Specify the base hostname
- Specify the number of hosts
- Specify first number to append to the base hostname
- Specify the "padding"
- Specify the starting IP
- Specify the subnet mask
- Specify the default gateway

Define OSCAR Clients		
Image Name:	oscarimage	Help
Domain Name:	cbi.utsa.edu	Help
Base Name:	oscarnode	Help
Number of Hosts:	0	Help
Starting Number:	1	Help
Padding:	0	Help
Starting IP:	129.115.16.1	Help
Subnet Mask:	255.255.255.0	Help
Default Gateway:	129.115.16.24	Help
Reset	Add Clients	Close

NOTE: You may only define 254 clients at a time!

Setup OSCAR Networking

		Setup Networking		
	it will appear in the left colu	MAC Address collection. When a new MAC address is received on the network, it will appear in the left column. To assign that MAC address to a machine highlight the address and the machine and click 'Assign MAC to Node'. Not Listening to Network. Click "Start Collecting MACs" to start.		
	, .			
	00:0C:29:9D:1D:71	□ All Clients □ -oscarnode001.oscar. □ -eth0 mac = □ -eth0 ip = 192.168		
 Collect MAC Addresses 				
 Optionally tweak SI installation mode 	Remove Remove All	j		
	Þ	MAC Address Management		
Build Boot CD	Start Collecting MACs	Assign all MACs	Assign MAC to Node	
	Delete MAC from Node	Import MACs from -	Export MACs to file	
OR	Insta	Installation Mode and DHCP Setup		
	systemimager-rsync -	Enable Install Mode		
 Setup Network Boot 	Dynamic DHCP update	Configure DHCP Server	ĺ	
Setup Network Boot	Boot Env	Boot Environment (CD or PXE-boot) Setup		
 Optionally choose to Use 	_ Enable UYOK	Build AutoInstall CD	Setup Network Boot	
Your Own Kernel (UYOK)		Close		

▲□▶▲圖▶▲≣▶▲≣▶ ≣ めぬの

Cluster Management	Cluster Management
Common Management Tools	Common Management Tools
LOSCAR	LOSCAR

Finishing Up!

Really, Its *that* simple!

- Go to "Monitor Cluster Deployment" to monitor the progress of the installation.
- Reboot the compute nodes.
- ► Go to "Complete Cluster Setup"
- Run the OSCAR Test suite (unless you're feeling brave!)
- Enjoy your new cluster!

- ► OSCAR comes with quite a few "standard" cluster packages.
- OSCAR uses SystemImager
- ► SystemImager is Good TM
- RPM packages may be added by placing them in the appropriate directory, rebuilding the image, and rebooting the nodes.

ROCKS Information

Vital Statistics:		
Version:	5.0	
Date:	November 12, 2006	
New development:	September 2008	
Distribution Formats:	tar.gz	
URL:	http://oscar.openclustergroup.org/	

Cluster Management Common Management Tools

ROCKS cluster distribution features:

- Supports X86, X86_64 processors
- Supports Ethernet networks
- Supports Specialized networks and components (Myrinet, Infiniband, nVidia GPU)

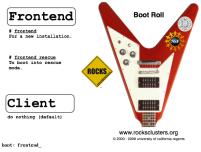
	▲日> ▲面> ▲回> ▲回> 回 ろんぐ	・ロト (四)・(目)・ 目、 しくの
Cluster Management Common Management Tools LROCKS	Cluster Management Common Management Tools Uther Popular Cluster Management tools	

Beginning the ROCKS Installation

For the Installation, you will need:

- ► Kernel/Boot Roll CD
- Base Roll CD
- Web Server Roll CD
- OS Roll CD Disk 1
- ► OS Roll CD Disk 2 OR
- ► ALL Red Hat Enterprise Linux 5 update CDs
- ► ALL CentOS 5 update 1 CDs
- ► ALL Scientific Linux 5 update 1 CDs

- ▶ Boot the "Kernel/Boot Roll CD" on the server
- You should see:



► Type "front-end" to begin the installation

Other Popular Cluster Management tools

- Xcat
- openMosix (RIP March 1, 2008)
- LinuxPMI Continuation of 2.6 branch of openMosix (*NOT* Single System Image)
- OpenSSI
- Scyld
- ► IBM's CSM
- ► Also notable: Sandia's CIT²

²It may not be the most popular, but it is well designed and pretty darn cool!

590

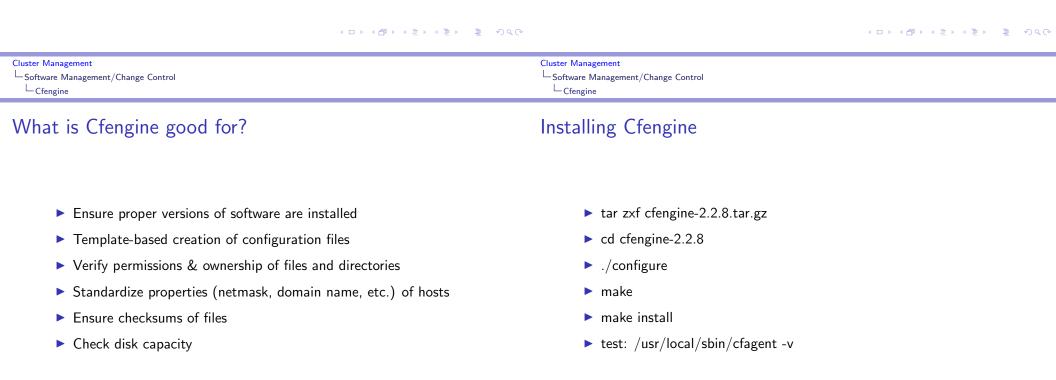
Cluster Management - Software Management/Change Control - Cfengine

What is "Change Control"?

Cfengine Information

- Automatically manage configuration files
- Take care of maintenance tasks like running backups
- Manage things like "cron jobs" in a centralized place.
- ... Automate and reduce the headache of administration!

Vital Statistics:		
Version:	2.2.8	
Date:	August 5, 2008	
Distribution Formats:	tar.gz	
URL:	http://www.cfengine.org/	



Getting Started with Cfengine

In order to get started with Cfengine, we will need 3 things:

- A crontab entry to run cfexecd periodically³
 0 * * * * /usr/local/sbin/cfexecd -F
- ► An update.conf file

► A cfagent.conf file

Cluster Management Software Management/Change Control Getting Started with Cfengine

update.conf — control section

control:

```
# distribute the files, then clean up our mess
workdir = ( /var/cfengine )
actionsequence = ( copy tidy )
policyhost = ( cfengine.hpc.unm.edu ) # master host
domain = ( hpc.unm.edu )
master_cfinput = ( /cfengine/inputs )
sysadmin = root@hpc.unm.edu
```

³Cfengine can also be run as a daemon. ← □ → ← ₱ → ← ₱ → ← ₱ → → ■ → ᠀ < ...

Cluster Management	Cluster Management
└─Software Management/Change Control	Software Management/Change Control
Getting Started with Cfengine	Getting Started with Cfengine

cfagent.conf — control section

cfagent.conf — files and directories section

control: domain = (hpc.unm.edu) netmask = (255.255.252.0)sysadm = (root@hpc.unm.edu) timezone = (MST) actionsequence = (mountall # mount filesystems in /etc/fstab # check the network interface netconfig resolve # check the DNS resolver # ''tidy'' Cfengine logfiles tidy # check file permissions files # ensure directories exist directories processes) # check processes

files:			
/etc/passwd	mode=644 ow	ner=root a	ction=fixall
/etc/shadow	mode=600 ow	ner=root a	ction=fixall
/var/spool/torque/pbs_environ	ment mode=64	4 owner=ro	ot action=fixall
/var/spool/torque/server_name	mode=644 ow	ner=root a	ction=fixall
#check that TORQUE directories	exist		
directories:			
/var/spool/torque/	owner=root	mode=755	action=fixall
/var/spool/torque/aux/	owner=root	mode=755	action=fixall
/var/spool/torque/mom_logs/	owner=root	mode=755	action=fixall

(etc.)

check important files

cfagent.conf — processes section

Here we define processes we want to ensure are running

- # We could also define ones we wanted to kill or restart
- # Strings are regular expressions used to match the name
- # of the process

processes:

"pbs_server"	matches=1	# ensure PBS is running
"maui"	matches = 1	# ensure Maui is running

Popular Parallel Shells

- ► PDSH
- Dancer's DSH
- Clusterit
- C3 tools

Cluster Management Parallel Shell Tools / Basic Cluster Scripting PDSH
Cluster Management
Parallel Shell Tools / Basic Cluster Scripting
PDSH

PDSH Information

PDSH Remote command modules

	Vital Statistics:
Version:	2.16
Date:	April 3, 2008
"Parallelism" :	"sliding window" parallel algorithm
Language:	C
Distribution Formats:	RPM, tar.gz
URL:	https://computing.llnl.gov/linux/
	pdsh.html

These are ways of accessing the remote nodes. Tune as per your security/performance requirements!

- RSH
- SSH
- Kerberos
- MRSH, QSH, MQSH, XCPU (whatever those are ;)

PDSH Node Specification

Cluster Management Parallel Shell Tools / Basic Cluster Scripting PDSH

PDSH Node Specification (cont.)

- Specify a list of hosts: pdsh -w node01,node05,node17 -- command
- specify a range of hosts:
 pdsh -w node01-node100 -- command
- Specify a range of hosts, excluding a set in the middle: pdsh -w node01-node100 -x node20-node30 -- command

- Specify a nodes in a netgroup "netgroup": pdsh -g netgroup -- command
- Exclude nodes in the netgroup "netgroup": pdsh -X netgroup -- command
- Execute a command on all nodes in a file: export WCOLL=/path/to/node-file pdsh -- command

	▲日▼▲ඛ▼★№▼★ № ● ●の人で	・ロット語・ (川)・ 加ァ かんの
Cluster Management	Cluster Management	
Dancer's DSH		

Dancer's DSH Information

Dancer's DSH Node Specification

Vital Statistics:		
Version:	0.25.9	
Date:	August 15, 2007	
"Parallelism" :	"Hierarchical invocation technique"	
	"4 nodes accessing 4 nodes"	
Language:	C	
Distribution Formats:	DEB, .tar.gz	
URL:	http://www.netfort.gr.jp/~dancer/	
	software/dsh.html.en	

- Use the global nodes file, /etc/dsh/machines.list: dsh -a -c -- command
- Use the list of nodes for "Rack 1" stored in \$HOME.dsh/group/rack1 dsh -g rack1 -c -- command

Clusterit Information

	Vital Statistics:
Version: 2.5	
Date:	August 15, 2007
"Parallelism" :	N-way Fanout
Language:	C
Distribution Formats:	.tar.gz
URL:	http://clusterit.sourceforge.net/

Cluster Management Parallel Shell Tools / Basic Cluster Scripting Clusterit

Clusterit Node Specification (Groups and Lumps)

- Groups are sets of nodes:
- GROUP:compute node01 node02
- Lumps are sets of groups:
- LUMP:cluster compute storage admin

	▲□ > ▲面 > ▲目 > 目 やくで	《□》《國》《토》《토》 토 원숙(
Cluster Management └─Parallel Shell Tools / Basic Cluster Scripting └─Clusterit	Cluster Management Parallel Shell Tools / Basic Cluster Scripting C3 tools (cexec)	

Clusterit Node Specification

C3 Information

- Specify a list of hosts: dsh -w node01,node04,node23 -- command
- Exclude a list of hosts: dsh -x node03,node09,node17 -- command
- Specify a group of hosts: export CLUSTER=/path/to/nodefile dsh -g compute -- command
- Specify a lump of hosts: export CLUSTER=/path/to/nodefile dsh -g cluster -- command

	Vital Statistics:
Version:	4.0.1
Date:	July 15, 2003
"Parallelism":	"Sub-Cluster Staging"
Language:	Python
Distribution Formats:	RPM, .tar.gz
URL:	http://www.csm.ornl.gov/torc/C3/
	C3softwarepage.shtml

Cluster Management
Parallel Shell Tools / Basic Cluster Scripting
C3 tools (cexec)

C3 Cluster Node Specification file format /etc/c3.conf

- Specify a cluster with a head node with an external interface named "external-name" and an internal interface named "node0" and 64 compute nodes named node01-node64.
- /etc/c3.conf contents: cluster my-cluster { external-name:node0 #head node node[1-64] #compute nodes }

Cluster Management Parallel Shell Tools / Basic Cluster Scripting C3 tools (cexec)

C3 Node Specification

- Specify the default cluster: cexec command
- Specify a subset of nodes in the default cluster: cexec :6-53 command
- Specify a list of clusters: cexec cluster1: cluster2: command

	< ロ > < 酉 > < 喜 > 、 言 う Q (?)	(ロ) (四) (四) (日) (日) (日) (日) (日) (日) (日) (日) (日) (日
Cluster Management └─Parallel Shell Tools / Basic Cluster Scripting └─Basic Cluster Scripting	Cluster Management Parallel Shell Tools / Basic Cluster Scripting Basic Cluster Scripting	
Basic Cluster Scripting grep is your (best) friend	More Basic Cluster Scripting awk is a pretty good friend too!	

- Find the CPU count on all of the nodes: pdsh "cat /proc/cpuinfo | grep processor | wc -l"
- Find nodes with the wrong image version: export VER="1.2.3" pdsh "cat /etc/image_version | grep \ "^\$VER\\$\" || hostname"

- Find nodes where the load is greater than 2: pdsh uptime | awk '{if(\$11 > 2.0){print}}'
- Find bad GM counts on all nodes: pdsh "/opt/mx/bin/mx_counters | awk '/bad/ {if (\\$2 > 0) {print;}}' "

Backup anything you can't recreate

Backup anything you can recreate but can't recreate quickly

- Use backup anytime it would take longer to rebuild and reconfigure than to restore.
- "Longer" may be in terms of staff time or elapsed time or both.
- ► Consider:
 - User directories (not scratch!)
 - Libraries and applications you've built on site
 - Tcl module files in /usr/share/modules/modulefiles/
 - System configuration files DNS, DHCP, NIS, etc. (Should that be everything in /etc/?)
 - Node images

Thanks to Roy Heimbach for contributing this slide!

Logging/Automated Log Analysis Tools:

SEC

- ► Logsurfer+
- 🕨 splunk

	く日 > 《加 > 《加 > 《加 > 《加 > 《四 > 《	<ロ> < 四> < 回> < 回> < 回> < 回> < 回> < 回> < 回
Cluster Management Logging/ Automated Log Analysis	Cluster Management Logging/ Automated Log Analysis Regular Expression Review	

What can we find in our logfiles?

What are we happily ignoring?

Regular Expression Review

Is that line noise?

- Evidence of misconfigurations:
 e.g. "/var/log/lastlog does not exist"
- Security violations
 e.g. Illegal users
- ► Hardware/Software errors e.g. Disk failures

This is a quick review of Perl Regular Expressions.

- Simple 'as-is' text string matching:
- "cat" or "dog"
- Meta-characters:
- ► {}[]()^\$.|*+?\

Cluster Management
Logging/ Automated Log Analysis
Regular Expression Meta-characters

Regular Expression Meta-characters

- ▶ . matches any single character
- * match the previous thing 0 or more times
- \blacktriangleright + match the previous thing 1 or more times
- ? match the previous thing 1 or 0 times
- ^ matches the beginning of the line
- \$ matches the end of the line
- ▶ \'escapes' the next character

Language:

URL:

Distribution Formats:

[] specifies a set or range of characters:
 eg. [a-z,A-Z,0-9] would match all alphanumeric characters

Cluster Management Logging/ Automated Log Analysis Regular Expression Meta-characters (cont.)

Regular Expression Meta-characters (cont.)

- ▶ {n} match the previous thing exactly "n" times
- ▶ {n,} match the previous thing at least "n" times
- {n,m} match the previous thing at least "n" times, but not more than "m" times
- () specifies groups of things or things to "save" the first group will be saved in \$1, the second in \$2, etc.
- | specifies "OR" inside of a group eg. (cat|dog) would match either "cat" or "dog"

	《曰》《國》《臣》《臣》 [] [] 《)	<ロ> < 部> < き> < き> きのへの
Cluster Management └─Logging/ Automated Log Analysis └─SEC		Cluster Management └─ Logging/ Automated Log Analysis └─ SEC	
SEC Information		<i>Quick</i> intro to SEC: SEC Components	
Version: Date:	Vital Statistics: 2.4.2 February 1, 2008	 Messages Single lines of text in a logfile 	

- Rules Do something in response to an incoming Message
- Contexts
 Passive structures to store Messages

.tar.gz, DEB, RPM, FreeBSD and OpenBSD

http://www.estpak.ee/~risto/sec/

ports, Gentoo portage

Perl

Cluster Management	Cluster Management
Logging/ Automated Log Analysis	Logging/ Automated Log Analysis
SEC	LSEC
Default SEC Rule	SEC Filtering Rule
Match all messages and print them	Ignore messages we're expecting
<pre># Print all messages type=single ptype=regexp pattern=.+ desc=unmatched message: \$0 # note \$0 is the entire message action=logonly This, or something like it, should be the last rule in your ruleset</pre>	<pre># This machine has 4 processors # Ignore messages reporting what we expect! type=single ptype=RegExp pattern=kernel: Total of 4 processors activated desc=correct processors initialized action=none</pre>
<ロ> < 昂 > < 言 > 、 言 ・ うへで	イロトイ団トイミト モーラ つので
Cluster Management	Cluster Management
イロトイラトイミト ミーシスで	<ロト < 部ト < 書ト < 書ト 書 つくで
Cluster Management	Cluster Management
Logging/ Automated Log Analysis	└ Logging/ Automated Log Analysis
LSEC	└ SEC
Cluster Management	Cluster Management └Logging/ Automated Log Analysis
Cluster Management	Cluster Management
Logging/ Automated Log Analysis	Logging/ Automated Log Analysis
LSEC	SEC

Cluster Management

Cluster Management

Logsurfer+ Information:

Vital Statistics:			
Version:	1.7		
Date:	December 2006		
Language:	С		
Distribution Formats:	.tar.gz		
URL:	http://www.crypt.gen.nz/logsurfer/		

Cluster Management

System and Cluster Security! Watch Out!

- Identify the Problem
- Security Strategies
- Dealing with Weaknesses
- Cluster Network Topologies
- Cluster Specific Issues
- Linux Tricks
- Checking Your Work

	・ロト 4 個 ト 4 差 ト 4 差 ト 差 りへで		< □ ▶ <	▶ ★ 豊 ▶	111	୬୯୯
Cluster Management └─ Security plans/procedures, Risk Analysis		Cluster Management — Security plans/procedures, Risk Analysis				
Define the Enemy		Attack Vectors				

- Data thieves
- Resource thieves
- ► Hackers there for various reasons
- Curies script kiddies
- Malicious script kiddies

Remote Attacks: Network Services allow access to the machine

 Local Attacks: Insecure Priveledged Binaries allow Priveledge escalation

Security Strategies

 \ldots besides cutting the wire

- Secure Communication
- Hunt and kill unneeded services
- Application configuration
- Protective Mechanisms

Cluster Management

Identifying Weaknesses

The key here is to strike a balance between security and useability

- Identify and categorize running services Are they *Really* needed?
- Identify sensitive information Passwords, Data, etc.
- Identify protective mechanisms TCPwrappers, iptables, firewall, etc.

	।□> ।∄> ।ই> ।ই> ।ই - ୬୧୯		- ロ > ・ (四 > ・ (三 > ・ (三 > ・)) への >
Cluster Management — Security plans/procedures, Risk Analysis		Cluster Management — Security plans/procedures, Risk Analysis	
Limiting Weaknesses		Finding services They <i>can't</i> hide!	
Local weaknesses:			
 Limit use of installed privledged binaries 		inetd(8) and xinetd(8) configuration files	
Removed setuid/setgid bits		chkconfig(8)	
If you don't use it, get rid of it!		init(8) scripts	
Remote weaknesses:		▶ ps(1)	
 Close unused ports 		► lsof(8) -i	
Limit access to ports		▶ nmap(1)	
If you don't use it, get rid of it!			

Killing Services

- ▶ kill(1)
- chkconfig(8)
- init(8) scripts
- inetd(8) and xinetd(8) configuration files
- chmod(1)

Common Cluster Services

- Login Service(s)
- File Transfer Service(s)
- File Service(s)
- ► Time Service
- Domain name service (DNS)
- Common Configuration Serices
 - DHCP
 - NIS
 - or ► LDAP

 - etc.

	・ロ・ ・聞・ ・言・ ・言・ うくぐ	・ロ・ (部・ (部・ (部・ 部・ 部・)のへる
Cluster Management	Cluster Management	
	Cluster Management	

└─Security plans/procedures, Risk Analysis

Security plans/procedures, Risk Analysis

Login Services

SSH Key Setup

rlogin, telnet, etc.

- SSH
 - Kerberized versions available
 - PKI (GSI) versions available

ssh-keygen -N "" -f /tmp/key
if you want password-less access
cp --force /tmp/key /root/.ssh/identity
rm --force /tmp/key
cat /tmp/key.pub >>/nfs/shared/authorized_keys
pdsh "cp /nfs/shared/authorized_keys /root/.ssh/"

Secure File Transfer

▶ scp(1)

- Encrypted connections
- Kerberized versions available
- ▶ Uses ssh(1)
- sftp(1)
 - "Simular" to ftp(1)
 - Encrypted connections
 - Kerberized versions available
 - Uses ssh(1)
 - Clumsy!

Secure X11 Connections

- Use ssh to "tunnel" X11 connections safely
- default ssh configuration files disable this
- ► To enable "X11 Forwarding":
 - In sshd_config add:
 - X11Forwarding yes
 - In ssh_config add: ForwardAgent yes ForwardX11 yes

ロット 4 聞 > 4 画 > 一 画 - うらる	・ロ・・西・・ヨ・・ヨー シック
---------------------------	------------------

Cluster Management	Cluster Management
Security plans/procedures, Risk Analysis	└─Security plans/procedures, Risk Analysis
	-Network Topologies and Packet Filtering

Using my admin tools from home...

SSH tunnels for the win!

- EVERYONE has used an X11 tunnel over SSH
- ► Have you ever forwarded something else?
- > Run administration tools from "inside" the firewall, but still at home
- ► Forward arbitrary ports Encrypted!
- ssh -v -L local-port:remote-machine:remote-port local-machine -l root
- ▶ ssh -v -L 1178:service1:1178 pq-admin.alliance.unm.edu -l root

Network Topologies and Packet Filtering

- Public Network Topology VS.
- Private Network Toplogy

Network Topologies and Packet Filtering

Public Network Topology

The easy way...

- Simpler to set up
- Allows direct access to compute nodes
- ► Worse overall cluster security
- ► ALL nodes need packet filtering, security tweaks
- All nodes are potential targets
- Better network throughput

Cluster Management

└─Security plans/procedures, Risk Analysis └─Network Topologies and Packet Filtering

Private Network Topology

Might be worth the extra headache

- Better security for entire cluster
- Relaxed security on compute nodes
- Only login/admin nodes on public network
- Compute/storage nodes access outside network via NAT
- Difficult to allow outside access to compute nodes

ふりん 前 《川を《川を《四を《日を

Cluster Management	Cluster Management
– Security plans/procedures, Risk Analysis	Lecurity plans/procedures, Risk Analysis
└─Network Topologies and Packet Filtering	-Network Topologies and Packet Filtering

Packet Filtering

Stateless:

Each packet is handled individually ipchains — (OLD!!! NOBODY uses this anymore!)

Stateful:

Each packet is viewed as a part of a session iptables — Modern, *probably* in your kernel.

- You can filter based on:
 - Network interface
 - Protocol type
 - Source address and port
 - Destination address and port
 - Other parameters depending upon the protocol

Keeps track of active connections

Stateful Packet Filtering

- Examines each packet based on their context
- Can provide a more useable system
- Controlled by iptables on Linux

Cluster Management

Protecting a single machine with IPtables

We're *not* doing NAT

iptables -A INPUT -m state ESTABLISHED, RELATED -j ACCEPT

- iptables -A INPUT -p tcp --destination-port ssh -j ACCEPT
- ▶ iptables -A INPUT -j REJECT

Protecting a network with IPtables

Hiding your cluster behind a NAT

- iptables -A INPUT -p tcp --destination-port ssh -j ACCEPT
- iptables -A INPUT -m state --state ESTABLISHED, RELATED -j ACCEPT
- iptables -A INPUT -i INTERNAL_INTERFACE -m state -state NEW -j ACCEPT
- ▶ iptables -A INPUT -j REJECT
- ▶ iptables -A FORWARD -j REJECT

	▲□> ▲□> ▲目> ▲目> 目 ろくで	<ロト 4 団ト 4 団ト 4 団ト 4 団ト 目 うん()			
Cluster Management Linux Tricks	Cluster Management				

/proc Protections

Turning on network stack security features

- Prevent address spoofing:
 - echo 0 > /proc/sys/net/ipv4/conf/*/accept_source_route
 - echo 1 > /proc/sys/net/ipv4/conf/*/rp_filter
 - echo 1 > /proc/sys/net/ipv4/conf/*/log_martians
- Disable ICMP redirects echo 0 > /proc/sys/net/ipv4/conf/*/accept_redirects
- Turn off bootp packet relaying echo 0 > /proc/sys/net/ipv4/conf/*/bootp_relay
- Ignore ICMP bad error responses echo 1 > /proc/sys/net/ipv4/icmp_ignore_bogus_error_responses
- Enable syncookie protection echo 1 > /proc/sys/net/ipv4/tcp_syncookies

Cluster-specific issues

Cluster-specific issues

- System backdoors:
 - cron
 - at
- One user per node guarantee
- Passwordless authentication

Cluster Management
Linux Tricks
Cluster-specific issues

One user per node

... or the right number of users per node

- Compute nodes should be wholly allocated to the user(s) that the scheduler has given them to
- Only the scheduler knows who owns the nodes
- ► Strategies:
 - Modify NIS maps
 - Modify /etc/passwd
 - PAM modules We (UNM HPC) use pam_pbssimpleauth distributed with TORQUE for most of our systems.

Cluster Management Linux Tricks Cluster-specific issues

Passwordless Authentication

- Job launch can't require passwords
- SSH can be used via RSAAuthentication (Public Key)
- Issues:
 - Management of host keys
 - Management of user keys

	< 日 > < 四 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <	(ロ) (四) (日) (日) (日) (日) (日) (日) (日) (日) (日) (日
Cluster Management	Cluster Management	
└─ Linux Tricks └─ Cluster-specific issues	Linux Tricks	

RSA vs. DSA (the low-down)

Checking Your Work

"In DSA, signature generation is faster than signature verification, whereas with the RSA algorithm, signature verification is very much faster than signature generation. ..."

(http://www.rsasecurity.com/rsalabs/faq/3-4-1.html) In a nutshell:

RSA can be used for both encryption and digital signatures. DSA is strictly a digital signature

- nmap port scanner
- Nessus vulnerability scanner
- Securityfocus.com
 - Search for your distribution & version
 - Compare vulnerabilities to services you run
 - Compare vulnerabilities to setuid/setgid binaries on your system
- Bugtraq for the seriously hardcore The up-and-coming info in the security world

Cluster Management	Cluster Management
	Linux Tricks
Checking Your Work	Checking Your Work

Finding listening services with lsof:

Finiding init.d started services:

lsof sl	nows	which	network	files	are	open:	
---------	------	-------	---------	-------	-----	-------	--

% lsof -i | awk '/LISTEN/ print \$1,\$(NF-2),\$(NF-1)' | sort | uniq condor_ma TCP service0.nano.alliance.unm.edu:1026 identd TCP *:auth inetd TCP *:ftp inetd TCP *:globus-gatekeeper inetd TCP *:gsiftp inetd TCP *:klogin

inetd TCP *:kshell
inetd TCP *:login

inetd TCP *:netsaint_remote

To find the services that will be started by default at the current runlevel using /etc/rc.d/init.d scripts: # chkconfig --list | grep 'grep :initdefault: /etc/inittab | awk -F: 'print \$2'':on | awk 'print \$1' | sort | column atd isdn random reconfig autofs keytable condorg sendmail netfs sshd crond network globus nfslock syslog

4	(日) 《四) 《川) 《山) 加 (山) 《四) 《日	・ロ・・雪・・声・ ゴー うくの
Cluster Management Linux Tricks Checking Your Work	Cluster Management	
Finding Network visible services	Regression Testing	

Nmap is your friend!

To find services visible from the network:

other-host# nmap host-to-be-looked-at

Port	State	Service
21/tcp	open	ftp
22/tcp	open	ssh
23/tcp	open	telnet
111/tcp	open	sunrpc
113/tcp	open	auth
513/tcp	open	login
514/tcp	open	shell
1026/tcp	open	nterm
4321/tcp	open	rwhoisw

Making sure stuff still works

gm

Your regression tests should:

Check your basic system components and tools

verifyd

pbs_mom

- Check your network(s)
- Check your important applications

Jim's Rule:⁴

If the cluster doesn't work for your users, the cluster *doesn't work*!

You're mostly on your own :P

... but its just some shell scripts...

- You can use tools like Cfengine to automate some of your regression testing
- > Your regression tests should be easy to run
- Your regression tests should produce a summary of successes and failures — a report at the end.
- Consider a suite of shell scripts
- Should the scripts attempt to repair any errors they find? (season to taste!)

System/Node/Software Change Management Logs

- Change management logs will save your backside!
- System administrators can be sloppy! :P :) Where did I put that??!
- Choose a tool that works well for the administrator(s) for the system in question.

(日)(日)(日)(日)(日)(日)(日)(日)(日)(日)(日)(日)(日)(· · · · · · · · · · · · · · · · · · ·
Cluster Management └─System / Node / Software Change Management Logs	Cluster Management How to know when to upgrade, trade-offs
Where to keep Change Management Logs?	How to know when to upgrade, trade—offs The Great Balancing Act!
Somewhere that you will actually keep them!	
A Wiki of some kind	 Security upgrades VITAL: if you have security concerns!
Emacs outline mode is nice!	VITAL: if you have ** A NETWORK CONNECTION! **
Really, whatever works for you and your staff!	Required features
I've seen sites alias editor commands in root's environment to require the admin to make a change management log when s/he	 Things needed to enhance the useability/stability of the system Software required by the users
edits a config file.	Tracking OS development
I won't tell if you're using a plain ASCII text file :)	You don't want to fall *too* far behind
 but if you do, please consider keeping it under some sort of version control :) 	 Upgrading several major versions is very painful! Keep your upgrades <i>relatively</i> small
	Latest development may not be what you want!

Cluster Management

Clumon Information:

V	Vital Statistics:				
Version:	2.0 Alpha				
Distribution Formats:	RPM,.tar.gz				
URL:	http://clumon.ncsa.uiuc.edu/				

Clumon

Bark +) Favorites Tools		- Favorites 💽 Media 😽	0 0. B R			
ddress 🙆 http:		Deardi M	Pavolices The Pavolices	Ø Ø. 3 m	• 🔜 🦔	Go Unis »	
NCSA PI	ATINUM CI	uster Monito	ər 👘				
Main						<u>^</u>	
Hosts							
Resources Queues			TT TTITT T	11111.1111			
Jobs							
Alerts Adm							
Notes							
Help							
Nodes: 518							
	Job Owner	r Job Name	Queue State	Nodes Time Used	1 % Time Allowed Max Time Allowed	i	
	32474 shuwang	g PYO1	standard RUNNING	32 06:28:03	8 54 12:00:00	1	
	<u>32502</u> dtoussai	i job2	standard RUNNING	80 05:19:21	I 44 12:00:00	j	
	<u>32516</u> degrand	p_4	standard RUNNING	64 04:57:50	0 41 12:00:00	1	
	<u>32517</u> schiu	DPPC400	standard RUNNING	6 16 04:58:17	7 41 12:00:00		
	<u>32518</u> degrand	pS_4	standard RUNNING	64 04:56:17	7 41 12:00:00	3	
	<u>32521</u> schiu	DPPC400	standard RUNNING	8 04:50:53	3 40 12:00:00	1	
	<u>32523</u> schiu	PORIN_S3	standard RUNNING	9 12 04:17:02	2 36 12:00:00	1	
	<u>32529</u> lek	SMUSH	standard RUNNING	6 15 03:47:20	0 32 11:55:00		
	<u>32535</u> kwthoma	as sp63-040912-36	standard RUNNING	9 13 01:46:15	5 15 12:00:00)	
	32536 pmarron	ie PT_14	standard RUNNING	8 01:47:04	4 15 12:00:00	1	
	and the second s		standard Queued	80	0 12:00:00	1	
	<u>32537</u> dtoussai	i job2	standard Greeded	00			
	10 - 10 - 10 - 10 - 10 - 10 - 10 - 10 -	i job2 K=0.01	standard Queued	8	0 12:00:00	1	

Cluster Management

Cluster Management

 cLUMSy

cLUMSy Information:

The Lightweight	Universal	Monitoring	System

... a work in progress ...

4 - 52 - 2 3 10 11 7 18 19 5 26 27 3 34 35 1 42 43 9 50 51 7 58 59	1 - 9 4 5 6 7 12 13 14 1 20 21 22 2 28 29 30 3 36 37 38 39 44 45 46 4	0 8 516 324 1 32 940 748 556 364	uster:po	bblano12:15:	28 - 6 - 0 1 2 3 4 9 10 11 1 17 18 19 20 25 26 27 20		Luster:hammer12:15:01 5 • 0 • 0 • 1 2 3 4 5 6 7 8 • 101112131415	10:50:90 1 2 3 4 5 9 1011121	678
3 74 <mark>75</mark> 1 82 83 9 90 91	76 77 <mark>78 79</mark> 84 85 86 87 92 93 94 95 - <u>busy</u> -	9 80 7 88 5 96	- <mark>down</mark>		33 34				
3 74 <mark>75</mark> 1 82 83 9 90 91 <u>1</u> - <u>free</u>	76 77 <mark>78 79</mark> 84 85 86 87 92 93 94 99 - <u>busy</u> -	980 788 596 offline		Req time	Used	Ea	rliest Start	Earlie	est Completion
3 74 <mark>75</mark> 1 82 83 9 90 91 <u>1</u> - <u>free</u>	76 77 <mark>78 79</mark> 84 85 86 87 92 93 94 99 - <u>busy</u> -	9 80 7 88 5 96 offline State				Ea -4:21:50:04	rliest Start Tue Sep 23 14:24:58		est Completion Tue Sep 30 06:24:58
374 <mark>75</mark> 18283 99091 1 - <u>free</u> cluster	76 77 78 79 84 85 86 8 92 93 94 99 - <u>busy</u> - job_id	9 80 7 88 5 96 offline State R			Used 117:49:28			1:18:09:56	
374 75 18283 99091 1 - <u>free</u> cluster	76 77 78 79 84 85 86 87 92 93 94 99 - busy - job_id 31556 31595	9 80 7 88 5 96 offline State R	Nodes 1	160:00:00 160:00:00	Used 117:49:28	-4:21:50:04	Tue Sep 23 14:24:58	1:18:09:56 4:14:56:44	Tue Sep 30 06:24:58
3 74 75 1 82 83 9 90 91 <u>l</u> - <u>free</u> cluster nano nano	76 77 78 79 84 85 86 8 92 93 94 9 - busy - job_id 31556 31595	980 788 596 offline R R R	Nodes 1 1	160:00:00 160:00:00	Used 117:49:28 49:02:53	-4:21:50:04 -2:01:03:16	Tue Sep 23 14:24:58 Fri Sep 26 11:11:47	1:18:09:56 4:14:56:44	Tue Sep 30 06:24:58 Fri Oct 3 03:11:47
3 74 75 1 82 83 9 90 91 <u>1</u> - <u>free</u> cluster nano nano nano	76 77 78 79 34 85 86 8 72 93 94 9 - busy - job_id 31556 31595 31679	9 80 7 88 5 96 offline R R R R	Nodes 1 1	160:00:00 160:00:00	Used 117:49:28 49:02:53	-4:21:50:04 -2:01:03:16	Tue Sep 23 14:24:58 Fri Sep 26 11:11:47	1:18:09:56 4:14:56:44	Tue Sep 30 06:24:58 Fri Oct 3 03:11:47

Vital S	Statistics:
Version:	0.0.0
Distribution Formats:	UNRELEASED Bug Jim

▲□▶ ▲□▶ ▲目▶ ▲目▶ 目 のへの

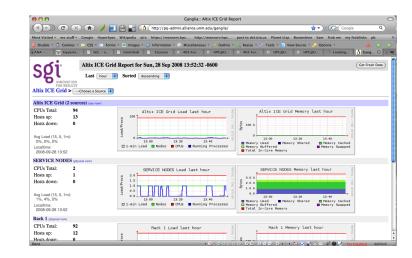
Cluster Management

Cluster Management

Ganglia Information:

Vital	Statistics:
Version:	3.1.1
Distribution Formats:	RPM,.tar.gz
URL:	http://ganglia.info/

Ganglia



◆□▶ ▲□▶ ▲□▶ ▲□▶ ▲□▶