

Beowulf Cluster Administration on Linux and HP-UX

Roger Goff Solutions Specialist

Hewlett-Packard Company America's Technical Computing Advanced Technology Center

<u>Roger Goff@hp.com</u> 3404 East Harmony Road Fort Collins, CO 80528 (970)898-4719 FAX (603)590-5337

Dr. Randy Splinter Solutions Specialist

Hewlett-Packard Company America's Technical Computing Advanced Technology Center

<u>Randy Splinter@hp.com</u> 238 Serenoa Drive Canton, GA 30114 (404)648-8003 FAX (678)493-8103

Topics to be covered:

- Beowulf Cluster Definition
- The physical layer how to connect to the compute nodes in a cluster, including a cost analysis of the approaches
- The logical layer how to administer and monitor a cluster

What is a Beowulf Cluster?

A kind of high performance, massively parallel computer, interconnected by a private, high-speed network. It consists of a cluster of PCs or workstations dedicated to running high-performance computing tasks. The nodes in the cluster don't sit on people's desks; they are dedicated to running cluster jobs. It is usually connected to the outside world through only a single node.

Beowulf Cluster vs. N.O.W.

A Beowulf cluster is distinguished from a *Network of Workstations* (N.O.W.) by several subtle but significant characteristics:

- The nodes in the cluster are dedicated to the cluster
- The cluster node interconnect is isolated from the external network, thus network load is determined by the application being run.
- Clusters do not have authentication concerns between the processors in the system
- OS parameters may be tuned to improve application performance without concern for impacts on general system performance



Beowulf Clusters

For more information

- http://www.beowulf.org
- http://www.beowulf-underground.org



Beowulf Cluster Administration - The Physical Layer

For the purposes of this discussion we are only considering accessing the individual compute nodes in the cluster beyond logging in to a node over the network. One may argue that if you can't telnet into a machine over the network you may as well physically recycle the power anyway, but it may not always be possible or desirable to take this approach.

Approaches to consider:

- Local, in computer room, access
- Remote access (beyond telnet)



Local Access

Use Keyboard, Video, Mouse (KVM) switch

- Pros:
 - * Up to 256 nodes can be administered from one console
 - Less expensive than providing remote administration capabilities for small numbers of nodes (less than ~40)
- Cons:
 - Must administer each system from within computer room
 - Cannot administer nodes over the network if network access to a node goes away
 - Can't access IPMI capabilities (more on this later)
 - Clusters limited to 256 node (without adding a second keyboard, monitor and mouse station for administration)



Remote Access

Uses an *administrative node* along with a terminal server

- Pros:
 - Can administer nodes from anywhere on the network even if the cluster or node's network is down (you must be able to reach the administrative node over the network)
 - One administrative node can access any number of cluster nodes
 - An administrative node can serve other purposes like file serving, installation services, cluster job management and running cluster management software
 - Can access IPMI capabilities (more on this later)
- Cons
 - Expensive solution for a small number of nodes (less than ~40) unless you can use an existing machine in your environment for the tasks
 - * System access speed limited to speed of serial terminal port

Intelligent Platform Management Interface

- IPMI was co-developed by Intel and HP to provide a low level consistent management interface for hardware vendors to provide with their systems
- IPMI support is built in to certain Intel motherboards which are used by many platform vendors today
- HP's NetServer products have IPMI support built in
- IPMI provides for the ability to remotely control:
 Device power control, hardware reset, remote bios console, power system monitoring, chassis intrusion monitoring, fan status monitoring and more.
- IPMI is implemented in hardware running on separate micro controllers, so monitoring does not impact a node's performance



Cost					
Comparison	Local Access Solution Cost	Remote Access Solution Cost			
16 nodes	\$ 900-1600 KVM switch <u>\$ 600 cables</u> \$1500-2200 total	\$3500 admin node \$ 144 Cat5 cable + connectors <u>\$1400 32 port terminal server</u> \$ 5,044 total			
32 nodes	\$1800-3200 KVM switch(s) <u>\$1260 cables</u> \$3060-4460 total	\$3500 admin node \$ 336 Cat5 cable + connectors <u>\$1400 32-port terminal server</u> \$ 5,236 total			
48 nodes	\$5400-9600 KVM switch(s) <u>\$2200 cables</u> \$7600-11800 total	<pre>\$3500 admin node \$ 544 Cat5 cable + connectors <u>\$2800 2x32-port terminal servers</u> \$ 6,844 total</pre>			
64 nodes	\$6000 KVM switch(s) \$4200 cables <u>\$2500 KVM monitor</u> \$12,700 total	<pre>\$3500 admin node \$ 760 Cat5 cable + connectors <u>\$2800 2x32-port terminal servers</u> \$7,060 total</pre>			

Assumption: racks of 2U compute nodes (20 per rack)



Cost of cabling is the key driver:

- KVM cabling costs
 - **♦** 6 foot \$35
 - ✤ 10 foot \$45
 - ✤ 15 foot \$55
 - ✤ 25 foot \$75
- Terminal servers use CAT5 or Telco cabling which is cheap
- Density of racks in computer room becomes a factor
- Providing remote access will be cheaper than local access for large clusters and when you already have a system that can perform the required administrative tasks

The Logical Layer - golden images

To ensure consistent results in computation and performance it is necessary that the software for each compute node in a cluster be configured identically. This is typically achieved by configuring and tuning one system to a known good state and then replicating that state to all other nodes in the cluster. We call the software clone of a known good state a "golden image."



Golden Images

The process for creating the actual image to be replicated differs from HP-UX to Linux in the tools that are available. The tool capabilities and modes of operation differ as well. In the next few slides we will look at HP-UX's Ignite-UX and the Linux tools *Kickstart* from Red Hat[®] and *SystemImager* from VA Linux Systems[®]



Ignite-UX for HP-UX

- Built by HP for HP-UX
- Can do:
 - * Attended installs via a GUI
 - Unattended package/depot based installs locally or across the net
 - ✤ Golden image installs from CD or across the net
 - Dynamic per client customizations based on client hardware
 - *bootsys* can take control of a client for either automatic or unattended installations
- Ignite-UX drawbacks:
 - ✤ Currently only runs on HP-UX

Ignite-UX for HP-UX

Image Creation Process

- Configure and test a client to use as your golden image source
- *make_sys_image* creates a golden image of the client that includes all customizations
- *archive_impact* creates volume size information for the file system volume groups of your golden image needed in the golden image configuration file on your Ignite-UX server
- Customize the Ignite-UX server configuration file for your golden image
- For more information:

http://www.software.hp.com/products/IUX/docs.html

Kickstart for Linux

- Built by and for Red Hat[®] Linux
- Provides unattended, across network installations, capable of repartitioning disk drives
- Essentially a scripted Red Hat[®] Linux install with post installation customization
- Red Hat[®] kickstart drawbacks:
 - Can't do true "golden image" based installs
 - Post install customization scripts can be difficult to debug
 - No other Linux distributions supported
 - * No GUI

Kickstart for Linux

Image Creation Process

- * Configure and test a client to use as your golden image source
- *mkkickstart* creates a kickstart configuration file that includes all configuration information for the golden image
- Create post install customizations in kickstart configuration file
- Create boot media from *boot.img* or *bootnet.img* images that you customize to invoke kickstart

For more information:

http://redhat.com/support/manuals/



SystemImager for Linux

- Built by VA Linux Systems
- Linux Distribution independent
- Provides unattended, across network installations, capable of repartitioning disk drives
- Golden image installs (not a package loader)
- SystemImager drawbacks:
 - Can't do package/depot based installs
 - * No GUI

SystemImager for Linux

Image creation process:

- Configure and test a "master client" system
- *prepareclient* run on master client collects disk partition information and configures *rsync* server on client system
- *getimage* run on "image server" to pull files from master client and create an auto installation script. Can also be used to "update" your golden image after changes are made to your master client
- *makeautoinstallcd* and *makeautoinstalldiskette* build boot media customized to access image server
- For more information

http://systemimager.org

Monitoring Tools

- Why monitoring tools?
 - Traditional methods of monitoring single systems will not work effectively
- Network Health Tools
- Node Health Tools
- Resource Monitoring Tools
- Caveats
 - I do not intend this to be an exhaustive overview
 - My goal is to cover enough of a variety of tools to give some flavor what is available

Network Health and Monitoring Tools

- * NetSaint (<u>http://www.netsaint.org</u>)
- Big Brother (<u>http://bb4.com</u>)
- * Ntop (<u>http://www.ntop.org</u>) or RedHat distribution CDs
- HP Network Node Manager



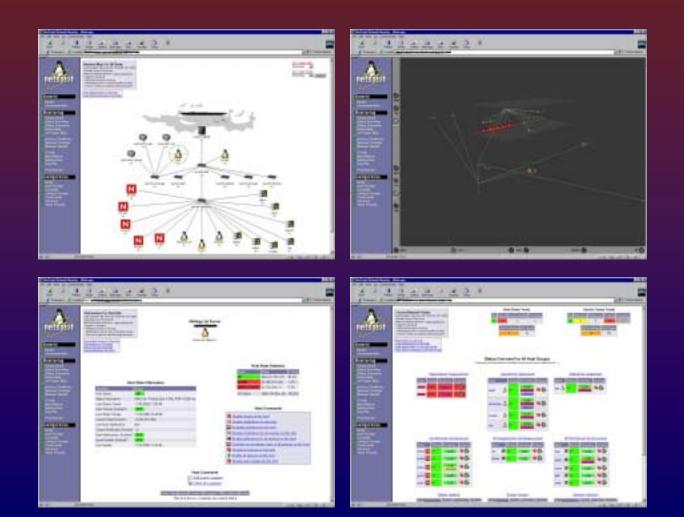
Net Saint

- Net Saint is a freeware tool downloadable from http://www.netsaint.org
 - Monitors hosts and services including disk, memory usage, processes, log files, etc
 - NetSaint can email or page when a problem shows up
 - Can automatically run scripts when a problem shows
 - ✤ Web interface available
 - Supports monitoring hosts





Net Saint Screenshots





Big Brother

- BB is a freeware tool downloadable from <u>http://www.bb4.com</u>
- ✤ BB is a web-based monitoring tool
 - * It requires the server run on a system that also has a web server running
- ✤ There is currently a
 - UNIX server and client
 - Linux server and client
 - NT client (not fully functional, not surprising?)
- The UNIX server and client compiled without problems on both HP-UX and RedHat 7.0
- BB is not very sophisticated, but gives you a simple quick view on the overall health of the cluster



Big Brother Screen Shots

✤ Main Screen

Netscap	pe: green	: Big Brot	ier - Status (🖗 Sun f	Feb 18 1	6,29,23 ESI	T 2001								_ = X
File Ed	It View	Ga C	o mini unicator												Help
₹ 🖋 B	ookmarks	👃 Locat	ion: http://	/Monet	ec/bb/b	b.htal						X	Cr.	What's Re	lated N
1	5	9	1 4	¥	a	D.		de.	a	0		-			
Back.	Eprura	rd Rela		ne -	Search	Netscape		Print	Security	Shop		Stop			
C CALIF	Admin	1 Sant	a Web Admi	nistratio	n Tanl		Lines	- Welcon	ie In Delt						
		* 0an		1111-0.00	1 1000	CONKIN	Links	neede	10 10 0 01	~					
	C L	ar 1 (j)								Sun F	eb 18 1	las 6:29:2	t upda 23 EST 2	ite 001
						die:					telast.				
			Monster												
			f3107ras												
			XCLASS												
		ok	e st	lention		(trouble		na repor	t 9	un av pilat	Die	effi <mark>e</mark>		networ	k ato
a												84	ile No	5 do 0	a 🍾



Network Connectivity History (past 24 hours)

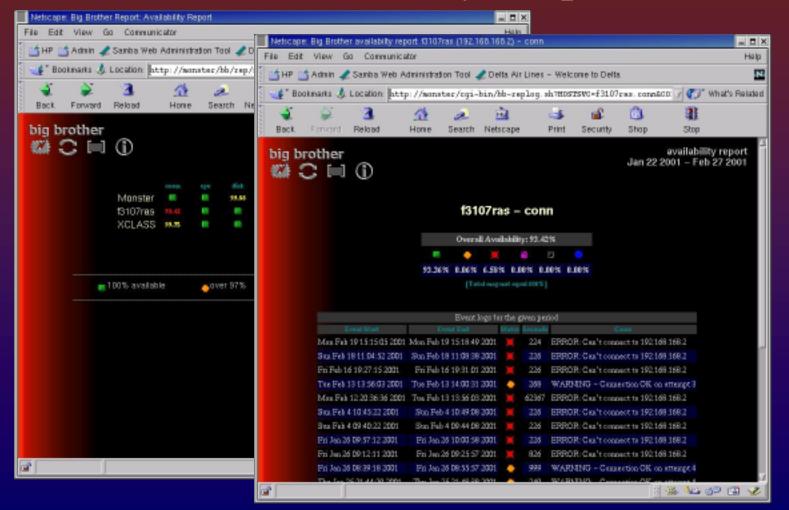
File Edit View Go Communicator Halp HP Admin Samba Web Administration Tool Detta Air Lines - Welcome to Detta Image: Communicator HP Admin Samba Web Administration Tool Detta Air Lines - Welcome to Detta Image: Communicator Image: Communicator <th>Netscape: Big Bro</th> <th>ther Statistica</th> <th>d Status: 101</th> <th>07ras (192</th> <th>2,166,10</th> <th>66.Z) -</th> <th>cont</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>2</th>	Netscape: Big Bro	ther Statistica	d Status: 101	07ras (192	2,166,10	66.Z) -	cont										2
Boolmants & Location attp://www.star/egi-bin/bb-hist.sh/HISTF/JLE-E3107res.com/EUT () What's Paladad Book Forward Reland Hone Search Netscape Print Security Shop Stop Dip Drother C C C C C C C C C C C C C C C C C C C	File Edit View	Ga Canni	unic ator													Hall	1
Bock Forward Reload Hone Search Netscape Print Security Shop Stop No Peb 26 17:39:38 EST 2001 Initiation	🕴 📺 HP 📑 Admin	🦧 Samba W	eb Administ	ration Too	10	elta Air	Line	s – Wel	come	to Deh	ta.					1	4
Back Forward Reload Hone Search Netscape Print Security Shop Stop big brother Distory Tue Feb 27 17:39:38 EST 2001 Inistory Tue Feb 27 17:39:38 EST 2001 Inistory Inistory </td <td>Bookmarks</td> <td>& Location</td> <td>http://wa</td> <td>natar/cg</td> <td>i-bin/</td> <td>/bb-hi:</td> <td>nt. ab</td> <td>178157</td> <td>FILE-</td> <td>63107</td> <td>z na. c</td> <td>onnes</td> <td>DIT y</td> <td>0</td> <td>" What's</td> <td>Relate</td> <td>d</td>	Bookmarks	& Location	http://wa	natar/cg	i-bin/	/bb-hi:	nt. ab	178157	FILE-	63107	z na. c	onnes	DIT y	0	" What's	Relate	d
Back Forward Reload Hone Search Netscape Print Security Shop Stop big brother Distory Tue Feb 27 17:39:38 EST 2001 Inistory Tue Feb 27 17:39:38 EST 2001 Inistory Inistory </td <td>1 4 4</td> <td>1</td> <td>45</td> <td></td> <td>8 11</td> <td>के</td> <td></td> <td>14</td> <td></td> <td>£</td> <td>13</td> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td></td>	1 4 4	1	45		8 11	के		14		£	13	1					
big brother Tue Feb 27 17:39:38 EST 2001 Non Feb 26 17:39:38 2001 Tue Feb 27 17:39:38 EST 2001 Non Feb 26 17:39:38 2001 Tue Feb 27 17:39:38 EST 2001 Sistery Tue Feb 27 17:39:38 2001 Sistery Sistery	Back Forward	d Reload								_	_	-		-			
Non Peb 26 17:35:38 2001 Tae Feb 27 17:38:38 ES1 2001 Non Peb 26 17:35:38 2001 Tae Feb 27 17:38:38 ES1 2001 State 7 4 Hours State 7 4 Hours Last 34 Hours State 7 4 Hours Last 34 Hours State 7 4 Hours Mon Feb 15 15:15:05:2001 State 8 (State 1) Last 34 Hours State 8 (State 1) Last 34 Hours State 8 (State 1) Last 134 Hours State 8 (State 1) Last 134 Hours State 8 (State 1) Last 14 Hours State 8 (State 1) Last 14 Hours State 8 (State 1) Last 14 Hours State 8 (State 1) Last 134 Hours State 8 (State 1) Last 14 Hours State 8 (State 1) Last 14 Hours State 8 (State 1) Last 14 Hours State 8 (State 1) Last 134 Hours State 8 (State 1) Last 134 Hours State 8 (State 1) Last 14 Hours State 8 (State 1) Last 134 Hours State 8 (State 1) Last 134 Hours State 8 (State 1) Mon Feb 15 11:13:131 Hours State 8 (State 1) Mon Feb 18 11:04:131 (State 2) 101:187	big brothe	ť															2
Main Feb 26 17:03:08 2001 Twe Feb 27 17:03:08 2001 15 13 20 21 22 20 1 2 4 5 7 5 9 11 12 14 15 16 17 Galor Feb 26 17:03:08 2001 Issue Feb 27 17:03:08 2001 Galor Feb 28 9 11 12 13 14 15 16 17 Galor Feb 28 Twe Feb 27 17:03:08 2001 Galor Feb 28 Twe Feb 27 17:03:08 2001 Galor Feb 28 Twe Feb 28 Twe Feb 28 Twe Feb 28 Galor Feb 28 Twe Feb 29:15:08 Twe Feb 29:15:08 Case 30 log envises (Fel HTML log) Twe Feb 19:15:15:02:001 Case 30 Non Feb 19:15:16:32:001 Case 30 State Feb 19:15:15:2001 Case 30 State Feb 19:15:15:2001 Case 30 State Feb 19:13:03:2001 Case 30 The Feb 19:13:03:2001 Case										1	fue Fe	eb 23	7 17:2	9:38	EST 2	001	Ī
15 15 23 21 22 23 0 1 2 3 6 7 5 9 10 11 12 13 14 15 16 17 f3107ras – conn Lest 24 Hours Image: Colspan=1 Colspa=1 Colspan=1 Col		Ψ															
15 15 21 22 23 0 1 2 3 4 5 6 7 5 9 10 11 12 13 14 15 15 17 f3107ras - conn Last 24 Hows												_					
f3107ras – conn Last 24 Howes Image: Ima														_			
Last 24 Hours Last 24 Hours Last 24 Hours Last 24 Hours Last 24 Hours Last 24 Hours Last 24 Hours Last 26 Hours on on on one on one one one one one one	15 19 20	11 22 23	0 1	2 2	4 3	56	7	5 5	20	- 11	12	12	14	15	16 1	7	
Lest 24 Hours Lest 24 Hours Lest 24 Hours Lest 24 Hours Lest 24 Hours Lest 26 Hours Le																	
Lest 24 Hours Lest 24 Hours Lest 24 Hours Lest 24 Hours Lest 24 Hours Lest 26 Hours Le					107		_										
Image:				15	51071	'as -	co	nn									
Loot 50 log ontrine (Full HTML log) East 50 log ontrine (Full HTML log) Rate Science Mon Ful 19 15:13:49 2001 Mon Ful 19 15:15:05 2001 Son Ful 18 11:04:52:2001 Son Ful 18 11:04:52:2001 Fit Ful 18 19:31:01:2001 Fit Ful 18 19:32:15:2001 Son Ful 18:13:01:001 Son Ful 18:13:01:2001 Son Ful 18:10:20:20:13:2001					Las	t 24 Hot	RIF.										
Inset 50 log centries [Fel HTM] log) Inset 50 log centries [Fel HTM] log) Non Feb 19 15:13:49 3001 Mon Feb 19 15:15:05 2001 Son Feb 18 11:03:38 3001 Son Feb 18 11:03:38 3001 Fri Feb 18 11:04:52:2001 Fri Feb 18 19:31:01 2001 Fri Feb 18 19:27:15:2001 Son Feb 19 31:01 2001 Son Feb 18 11:04:52:2001 Son Feb 18 11:04:52:2001 Son Feb 18 11:04:52:2001 Son Feb 13 14:00:31:000 Son Feb 13 14:00:31:0001						× .	Ð										
Laset 50 kog enticies (Fail HTbdl. kog) Man Feb 13 15:13:49 3001 Man Feb 13 15:13:49 3001 Man Feb 13 15:15:05 2001 Sun Feb 13 11:03:38 3001 Fei Feb 13 11:04:52 2001 Fei Feb 13 14:04:52 2001 Fei Feb 13 14:00:31 3001 Fei				1007	E 07E	076 078	E 016	076									ſ
Non Feb 19 15:18:49 2001 224 Mon Feb 19 15:15:05 2001 224 Son Feb 19 11:08:38 2001 101187 Son Feb 18 11:04:52:2001 226 Fri Feb 16 19:31:01 2001 1242431 Fri Feb 16 19:27:15:2001 226 Toe Feb 13:14:00:31:2001 226																	
Mon Feb 19 15:18:49 3001 Image: Control of Contro of Control of Contr																	
Mon Feb 19 15:18:49 234 Mon Feb 19 15:15:05:2001 234 Son Feb 18 11:08:38:2001 101:187 San Feb 18 11:04:52:2001 236 Fri Feb 16 19:37:15:2001 236 Toe Feb 13 14:03:31:2001 236 Toe Feb 13 14:03:31:2001 236				Loot St	log ent	ntere (Pr	d En	ML loci									
Mon Fab 19 15:15 05 2001 224 Son Fab 18 11:03:38 2001 101187 Son Fab 18 11:04:52 2001 226 Fit Fab 16 19:31:01 2001 142431 Fit Fab 16 19:27:15 2001 226 Too Feb 13 14:03:31 2001 226 Too Feb 13 14:03:31 2001 226					Date												
Sun Feb 18 11 03 38 2001 ID1187 Sun Feb 18 11 04 52 2001 226 Fit Feb 16 19 31 01 2001 142431 Fit Feb 18 19 27 15 2001 226 Toe Feb 13 14 00 31 2001 226 Toe Feb 13 14 00 31 2001 226																	
Stur Feb 18 11 04 52 2001 (226 Fri Feb 16 19 31 01 2001 (226) Fri Feb 16 19 27 15 2001 (226) Toe Feb 13 14 00 31 2001 (226) Toe Feb 13 14 00 31 2001 (226)								224									
Fil Feb 16 19 31 01 2001 142431 Fil Feb 16 19 27:15 2001 236 Toe Feb 13 14:00 31 3001 278804																	
Fei Fab 16 19 27:15 2001 💓 236 Toe Feb 13 14:00 31 3001 🖷 278804																	
Toe Feb 13 14:00 31 2001 💌 278804																	
7 7 1 4 3 4 5 5 6 5 5 3 3 4 3																	
🗃 🐘 😼 🜮 🖼 🤣				Toe Feb 1	3 14:00	31 2001		2768	04								1
	a											1	-14	10	32 0	3 2	

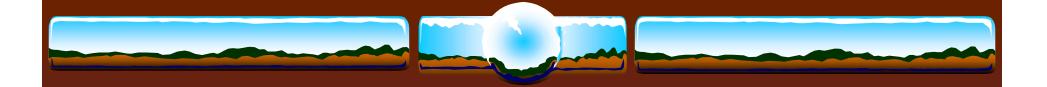
Network Connectivity Report

Netscape	s. Big Brothe	er Report: Ava	lahiity Re	sport								
File Edit	View G	ia Communi	cator									Halp
HP 🗃	SAdmin 🦼	Samba Web	Administr	ation To	o ol 🦧 D	elta Air I	Lines - W	elcome	to Delta			N
📲 Boo	okmarks 🞄	Location: ht	tp://wor	nter/	bb/zep/	rep-128	78.html				7 CT V	What's Related
1	2	3	1	3	2	È.	4	. 1	aî.	<u>s</u>	-	
Back.	Forward	Reload	Home	500	uch Ne	etscape	Prin	1 50	scurity	Shop	Stop	
	c 🗐	1								Jan 1:	availabilit 2000 – Feb	y report 27 2001
		Monster			99.85			99,09				
		t3107ras XCLASS	99.42 99.75			97.34		99,39		99,79		
		100% availab	le .	•aw	er 97%		eless that	1 97%	ыN	o stats for j	period	
											dov	beolev sow?
a .										100	*	P 🖬 🎸



Network Connectivity Report (Details)





Network Connectivity Report (text format)

	vertsc ap e	s Big I	Brother	Report Avai	ishiity Re	eport														
Fil	e Edit	Vie	w Ga	Connunio	cator		Netscare	Big Broth	er availabilty report	010Trat (197	165 165 2	Haln								
1000	THP 🖸	5 Ada	in 🥠	Samba Web	Administr	ation Tool			Go Communicator	Netscape.										
Part 1	€" Boo	itmari	s 🎄 L	ocation he	tp://wor	wter/bb	E COS HP C	Admin	🛃 Samba Web Admir	File Edit	View (Ga Cann	runicator							Halp
100	4	ŝ	ž	3	1	à			Location http://	2 🗂 HP 🚮	Admin 🚽	🖉 Samba Y	Yeb Administra	ton Taol .	n Delta Air Li	nes - Wel	come to De	ita.		N
5000	Back.	Forv	rand	Reload	Home	Searc		N-	Counter processor	📲 Boo	inaris 🤳	Location	http://www	iter/bb/s	ep/avail-f3	107cmm-c	mn-13019.	txt	7 🕐 V	mat's Related
	in he	-					C Reck	Francisco de	Dalaad Ua	1	2	3	1	a.	<u>1</u>	4	1	۵.		
	nig bi						3 Back	Forward	Reload Ho		Forward	Reload	Home	Search	Netscape	Print	Security	Shop	Stop	
		م		Ψ			big br	other		Presilability Jan 22 2001	· Pap 27	2001								
							- 12 A		1 1				f3107res - c							
				Monster		er i							Reallability							
				t3107ras	33.41	-						Red 6.58%	Sellow Gree 0.05% 55.3	n Furple	c Olear Bit	RE				
				XCLASS	99.75											~~				
										Event Stort			Event Logo f	or the gir		nter Secon	str Carrie			
										tion Feb 15 Sun Feb 18 Fri Feb 16	15:15:05 11:04:52 15:27:15	2001	tion Feb 15 1 Sun Feb 15 1 Fri Feb 16 1	5:10:45 20 1:05:35 20 5:31:01 20	001 1965 001 1965 001 1965	226	ERROR: ERROR:	0an't connec 0an't connec	t to 152.168.1 t to 152.168.1 t to 152.168.1	168.2 168.2
				00% availabi	le .	over				Toe Feb 13 tion Feb 12 Son Feb 4	13:56:05 0	2001	The Feb 13 1 For Feb 13 1 Son Feb 4 1	4:00:31 20 3:55:03 20	901 941 901 246	Low 268 62361	7 ERROR:	0 am't commerci	n OE on sttem t to 152.168.1 t to 152.168.1	168.2
										Frit Jan 26	05:40:22 2 05:57:12 2	2001.	Son Feb 4 0 Frit Jan 25 1	5:44:05 20 0:00:55 20	901. 202 901. 202	226	ERROR: ERROR:	Oan't connec Oan't connec	t to 152.168.1 t to 152.168.1	168.2 168.2
									Event Start	Fri Jen 26 Fri Jen 26 Thu Jen 25	08:35:18 2	2001	Pri Jen 26 0 Pri Jen 26 0 Tho Jen 25 2	5:55:57 20	001 OR.	Low 335 Low 245	WARKEN	- Connection	t to 152.168.1 n OE on attemp n OE on attemp	pt 4
									191515052001 Mo	Wed Jan 24 The Jan 25	05:55:27 2	2001	The Jan 25 2 Tox Jan 23 1	1:44:25 20	001 inté		Z ERROR:	Can't connec	t to 152.168. n OE on attemp	168.2
									1811.04:52 2001 So											
									16 19:27:15:2001 Fr 13 13:56:03:2001 To				Time Orithcom	L/OffLine	2 days 8 hot	ers 40 mins	45 3403			
									12 20 36 96 2001 To											
								Stu.Feb	4 10 45 22 2001 5											
									4 09 40:22 2001 S											
									6 09:57:12:3001 F											
6									16 D8:39 18 2001 F											
									9501.4400.9001 Th											
										a [í							3	*******	0 🖼 🍾
											_			_			_	10.00		



Ntop

- Ntop is a tool for displaying network usage
 - ✤ Has both a curses interface and a web interface
- Freely downloadable from <u>http://www.ntop.org</u>

]		Welcome to stop!		P
Welcome		Global Traffic St	atistics	
to to	Ne Interface Type			Ethemet (eth0)
ntop!	Local Domain Name			tecsiel it
	Sampling Since		Ri May 19 09:142	22 2080 [1:26:36]
About stop		Total		188,547
		Dropped by the kernel		0
About ntop		Dropped by ntop		0
Data Revd		Unicast	61.0%	115.090
		Broadcast	13.1%	24,665
All Protocols :		Multicast	25.9%	48,792
olP o Throughput v Stats o Muticast o[Traffic] o Hosts		halticest		
o Throughput v Stats o Multicast o Traffic o Traffic o Domain o Domain o Diugins b IP Traffic b IP Protocols		hiltscort Bromicest	Inicost	
O Throughput Stats O Multicast OTraffic OTraffic			Inicost	
O Throughput Stats Multicast ofTraffic oftouthouthouthouthouthouthouthouthouthout			Inicost	28 tytes

	intop 0.0.1	(May 19 2	2000) liste	ning on [hm		_
6606 Pkts/770.7 Kb	[IP 703.7	<u>Kb/Other 6</u>	57.1 Kb]	Thpt: 211	1.9 Kbps/349	
Host	Act		Sent	TCP	UDP	I CMP
more	В	257.4 Kb	281.9 КЬ	256.6 Kb	769	0
zetant	В	204.2 КЬ	232.3 КЬ	204.2 КЬ	0	0
tar	В	42.9 Kb	19.5 Kb	42.9 Kb	0	0
ibook	В	32.7 КЬ	4.7 Kb	32.7 Kb	0	0
tecserv	R	791	0	0	595	196
bugnoli	в	602	1.4 КЬ	0	602	0
urano	в	496	5.1 Kb	0	496	0
utlrouter	R	98	0	0	0	98
mis	S	0	212	0	0	0
fiorella	s s	0	486	0	0	0
piutltst02	S	0	1.4 КЬ	0	0	0
mostardi	S	0	952	0	0	0
193.43.104.55	S	0	588	0	0	0
itest1	S	0	928	0	0	0
rolly	S	0	46	0	0	0
itinŽ	S	0	92	0	0	0
3comhub1	S	0	610	0	0	0
re	S	0	5.6 Kb	0	0	0
pi100	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	0	1.2 КЬ	0	0	0
leardini	S	0	546	0	0	0
mbeng	S	0	602	0	0	0
i test2	S	0	600	0	0	0
fossati-a	S	0	960	0	0	0
hpwsutl	S	0	3.1 КЬ	0	0	0
catle	S	0	120	0	0	0
aut01b	S	0	243	0	0	0
biu	S	0	542	0	0	0
artico2	S	0	226	0	0	0

HP OpenView Network Node Manager

- NNM is part of the HP OpenView suite of products for enterprise management
- Supplies much of the same functionality as both Net Saint and Big Brother
- Probably not a good choice unless
 - The cluster is to be integrated into the enterprise environment
 - ✤ A well supported tool is required
 - The tool needs to be integrated into either OpenView or another enterprise management suite, such as Tivoli

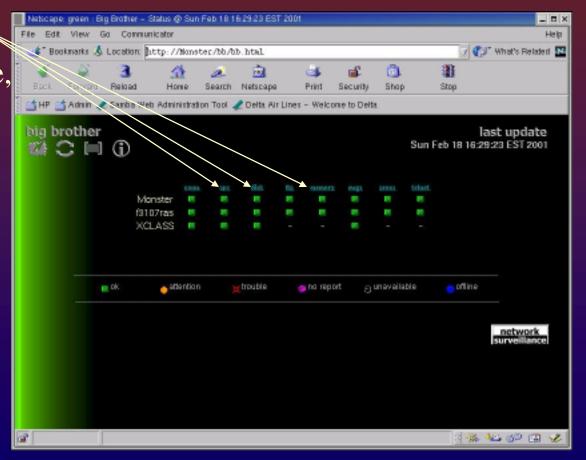


Node Health Tools

- * Net Saint (<u>http://www.netsaint.org</u>)
- Big Brother (<u>http://www.bb4.com</u>)
- Xosview (<u>http://lore.ece.utexas.edu/~bgrayson/xosview.html</u>) or RedHat distribution CDs
- ✤ HP Glance

Big Brother Node Health

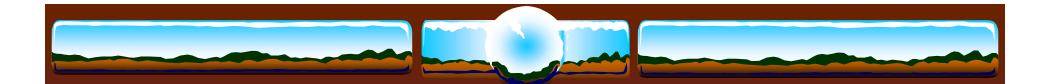
 Information on the CPU usage, disk usage, memory usage all can also be quickly seen from this screen





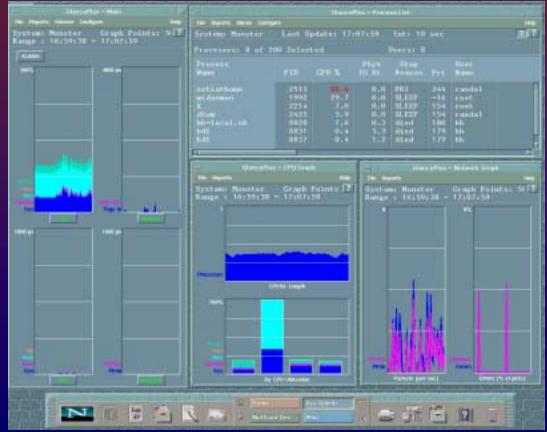
- Monitors and a health using a V usindary
- Monitors node health using a X-windows interface





Glance Screen Shot

HP Glance is an add-on product for HP-UX systems. It has a curses and Motif GUI





Resource Monitoring

Load Sharing Facility (LSF) (<u>http://www.platform.com</u>)
Linux Resource Manager (<u>http://www.arc.unm.edu/~lrm/</u>)
Maui Scheduler (<u>http://mauischeduler.sourceforge.net/</u>)
Portable Batch System (PBS) (<u>http://www.openpbs.org</u>)



What is LSF?

- ✤ A heterogeneous network of computers is viewed as a single system
- Users are no longer limited to the resources on their local workstations
- Programs do not need to be changed or rewritten for use with LSF
- Only a few simple commands are required to submit jobs and execute tasks
- * LSF automatically selects the best host to execute a job or task



The Complete LSF Suite of Products

- LSF Standard Edition
 - The Standard in Application Resource Management
- ✤ LSF Parallel
 - Application Resource Management for Parallel Computing
- ✤ LSF Make
 - Make processing for Heterogeneous Distributed Computing
- LSF JobScheduler
 - Job Scheduling for Heterogeneous Distributed Computing



The Complete LSF Suite of Products

- LSF MultiCluster
 - Application Resource Management for Enterprise Computing
- LSF Analyzer
 - Graphical Performance Analysis Tool for Application Resource Management
- LSF Client
 - Extending Application Resource Management to the Desktop

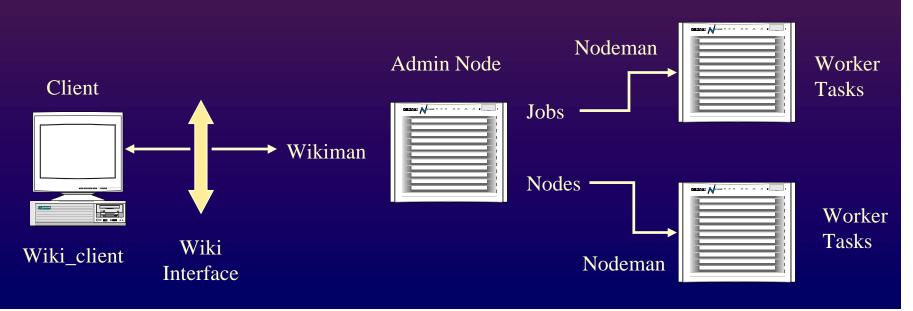
Linux Resource Manager

History Lesson

- The Maui scheduler was originally developed to be dependent on the IBM SP Load-Leveler API
- As interest in the Maui Scheduler for Linux, IRIX, HP-UX, and Windows NT grew it was necessary either to write the interfaces to existing Resource Managers, or to develop a Resource Manager (RM) specifically for the Maui Scheduler
- In 1998, a Maui High Performance Computing Center team started to develop a generic resource manager called Wiki, from what was the Wiki RM
- The Linux Resource Manager development began at the <u>Albuquerque High</u> <u>Performance Computing Center</u> and is known as the Linux Resource Manager

Linux Resource Manager

- Nodeman process runs on each compute node
- Wikiman, Maui Scheduler process runs on admin node
- Wikiman gathers state of compute nodes from Nodeman
- Wikiman communicates with Maui Scheduler
- Wikiman launches MPI job on the compute nodes



Compute Nodes

Linux Resource Manager

- *showq*, *shownode*, *showjob* are used to view queues, processes running on a node, and to view the resources consumed by a particular job
- Use submit job program to submit a job
- Use canceljob program to cancel a job



PBS

- Developed for NASA by Veridian Systems
 - Veridian maintains a professional version on a wide variety of platforms
 - More information at <u>http://www.pbspro.com/</u>
- xPBS graphical interface and command line interface
- Job priorities
- Job Interdependencies
- Cross-system scheduling
- Security and ACLs
- Job Accounting
- Automatic load-balancing
- Fully configurable
- ✤ Supports MPI, PVM, HPF
- Automatic file staging

Maui Scheduler

- Maui Scheduler is an advanced reservation based High Performance Computing batch scheduler
- Supported on SP, O2K, and Linux/UNIX clusters
- Can be used to extend the functionality and improve the efficiency of sites utilizing the PBS and Loadleveler batch systems



ServiceControl Manager

- Centralized
- Multi-system
- System deployment and recovery
- Multi-system software management
- System configuration management
- Hardware and software fault management
- For more details see the presentation by Humberto Sanchez titled "An Overview of the SCM and the New 2.0 Features"