

;login:

THE MAGAZINE OF USENIX & SAGE

August 2001 • Volume 26 • Number 5

inside:

CLUSTERS

AFFILIATED HEALTH SERVICES
INFORMATION SYSTEMS CLUSTERS

by Richard Schilling

Special Focus
Issue: Clustering

Guest Editor: Joseph L. Kaiser

USENIX & SAGE

The Advanced Computing Systems Association &
The System Administrators Guild

Affiliated Health Services information systems clusters

by Richard Schilling

Richard is the web integration programmer and webmaster at Affiliated Health Services, a hospital in Mount Vernon, Washington. He has a B.S. in computer-science and is currently an MBA graduate student. Richard's work with-clustering is focused on developing distributed and parallel algorithms for use in the health care industry.



RSchilling@affiliatedhealth.org

I have three clusters, though I would not classify them as true Beowulf-class machines yet. I'll describe them and the work in bridging them, their limitations, the lessons learned, and future plans.

Cluster 1: At Work

I am the Webmaster and Web integration programmer at Affiliated Health Services (<http://www.affiliatedhealth.org>). I have set up a cluster of three Pentium-based workstations that were taken from spares. The head node is a dual-named host which forms a link between the corporate network and cluster. The nodes of the cluster do not see the corporate network, but rather, the head node uses the cluster for extra file storage and processing capacity as needed. The purpose of the cluster is to establish performance minimums and a baseline equipment list for future development. I use the nodes in the cluster for file sharing. As more computers become available, I will make use of them also. In the near future, due to upgrades, I will have quite a few. As time permits, on this cluster, I benchmark various algorithms used in health care, such as decision support routines, medical billing functions, and data warehousing.

The basic approach to this is making the most of available tools while keeping the configuration as simple as possible. I'm not trying to break any speed records with this cluster but rather am focused on getting set up to solve basic business problems in health care (processing patient data) and then scaling up to heartier hardware.

Most of my time spent working on this cluster has been to solve normal business problems. Anything beyond that so far has been dedicated to trying out various tool sets and configurations. Hopefully, I'll soon get more into the serious benchmarking.

The nodes all have the same operating system, FreeBSD 4.1. Here is a list of the software I have on the nodes:

HEAD NODE

I use the head node to start jobs and, as time permits, test out "true" Beowulf-class applications. Day-to-day use has this computer doing more Web serving and data processing related to my work as a Webmaster, such as logfile crunching.

NFS; Apache: modules include mod_ssl, mod_perl, mod_php4; Open SSL; clusterit; CVS; PostgreSQL; MySQL; Kerberos; SSH (PVM uses this for remote logins); Kaffe; OSKit; PVM and jPVM; GNU tool kits (of course); Palm development tools (prc tool set); X Windows; grass (GIS); aero (modeling); Games: freeciv, xpilots, ACM.

NODE 1

This node is used primarily for extra disk space. It hosts the CVS repository and exports the /usr partition to the other nodes on cluster.

Software: Kerberos V; SSH; GNU tool kits.

NODE 2

This node is similar to Node 1 in that it supplies extra disk space and processing power when needed.

Software: Kerberos; SSH; GNU tool kits.

NETWORK

The following is used for networking equipment in the cluster: Intel Etherexpress PRO network cards; Bay Networks Baystack 102 hub.

Cluster 2: At Work

The second cluster we have at work is in the process of being implemented. It is our main information system, and although I don't know enough about it to know if it qualifies as a true Beowulf, it sure seems like one so far. We'll have about 25 nodes, and the software is a canned package created by MediTech (<http://www.meditech.com>). Mostly, this cluster does distributed processing, distributed batch processing, and fail-safe backup work, as opposed to running true parallel algorithms. The manufacturer should be able to verify that. This will be installed this year.

Cluster 3: At Home

As if I didn't have enough to do already . . .

Here's what the cluster at home looks like:

Sequent 2000 NUMA machine: head node (AT&T System V); two Pentium-class machines (FreeBSD 4.1); one DEC VAX 4000 (being set up; it will eventually run ULTRIX or NetBSD); one Power Macintosh (MkLinux and MacOS).

This cluster is a true experiment in heterogeneous computing environments. I have two late-model Pentium-class PCs, a Sequent 2000 NUMA machine, a Power Macintosh, and a DEC VAX 4000. Like the cluster at work, this one is being used to establish baseline minimums of hardware and software performance for present and future software projects. As opportunity arises and my software base becomes stabilized, I'll get more involved with the serious benchmarking. But this cluster also has another purpose: to reduce the steps of adding new heterogeneous nodes as much as possible. Ideally, once the operating system is installed on a machine, I'll simply plug in a new machine and start up the appropriate software.

Using the Power Macintosh is a curious choice, and I get some flak for it, but the fact is that it has turned out to be a nice little machine to use as a clustering proving ground. It has a PC daughter card in it, too, so I have a little extra PC power when I need it (while running MacOS). I also run MkLinux on this machine, which I use when I'm doing all the clustering work. Since Macintosh printers have almost always been Postscript printers, I have absolutely no problem with network printing from the other nodes on the cluster.

The two PCs mainly work as file servers at this point but will complement the Sequent and VAX boxes when I have those two set up. The Sequent machine is a recent addition and will work as the head node when it is fully configured (sometime this summer).

The focus for this cluster is to identify as many boundaries as possible in a heterogeneous system. The use of older equipment is a non-issue, since I'm sticking to the UNIX (variant) and GNU standards. Overall, this has done well, and I've even gotten the OSKit on one of the PCs to work, so I've made my first custom operating system there.

NETWORK

A common hub bought at the local computer store, any network card that works in the PCs; Mac, VAX, and Sequent machines all have network cards built in, so I don't have to worry about those.

This cluster is a true experiment in heterogeneous computing environments. I have two late-model Pentium-class PCs, a Sequent 2000 NUMA machine, a Power Macintosh, and a DEC VAX 4000

Clustering computing is currently available and affordable for the enterprise and home environments

Bridging the Work and Home Clusters

I use my work laptop to transfer data between the home and work clusters. If the opportunity arises in the future, I'll have the home cluster dial into the work cluster.

Limitations

Ironically, the limitations I am experiencing are not primarily due to software. The OS base for Linux and FreeBSD is very stable for what I'm doing. The primary limitation I have run into so far is finding the time to dedicate to setting up the hardware, adding hard drives, and creating network settings, like IP addresses. Oh, and then there's physical space – my garage is full of computers and spare parts.

The Lesson

Here is what these two clusters have allowed me to demonstrate: clustering computing is currently available and affordable for the enterprise and home environments. I'd be lost had I not received a computer science degree and previous exposure to the UNIX environment, but it's getting easier for non-tech heads all the time.

As a general rule, it's a very bad idea to invest in the latest and greatest hardware unless it actually improves your capabilities from a clustering/functional standpoint. The idea that new (proprietary) software products can stimulate PC sales is preposterous. Hardware should only be invested in when there is too much data to handle with the existing equipment, or the cost of new equipment drops low enough to make it cost effective. Even then, you should not have to change software – recompile, yes, but never change.

Future Plans

Here is a list of future plans I have for my clusters:

- Addition of 64-bit Sun Microsystems nodes
- Work on software, including a distributed version of PostgreSQL, 64-bit
- FreeBSD, MkLinux port for G4, general benchmarking, utility set for aerospace engineering applications, modeling, Java development,
- Windows emulation, and, of course, game development
- Porting of several open source packages for the Windows environment (to try and sway the Windows crowd to the open source arena)
- Robotics
- A variety of artificial intelligence work (my computer science specialization was AI)