



January 2003

HPC Times

IN THIS ISSUE:

- ▶ **News from Microway®**
 - ▼ Plans for 2003 [Pg. 1]
 - ▼ New Life Sciences Customers [Pg. 2]
- ▶ **You Can Count on It** by Stephen Fried, CTO and President
 - ▼ BPROC/Linux BIOS Research at LANL [Pg. 3]
- ▶ **Tech Notes** by Nina Nitroy, Technical Support Manager
 - ▼ Commercial Compiler Options for Clusters [Pg. 5]
 - ▼ Customizing Your Cluster [Pg. 6]
- ▶ **Parallel Thoughts** by Jay Owen, VP Business Development
 - ▼ Commercial vs. Open Source Applications - Tradeoffs [Pg. 7]
- ▶ **HPC Forum:** Creating a community of questions and answers.

Microway's Plans for 2003

Microway's managers spend time each January setting new goals for corporate growth. These include specifying new products, projecting R&D funding, and expanding our expertise into arenas which we believe will benefit our customers. Our business continues to be strong with customers projecting increased requirements for clusters and services in 2003.

In 2002, Microway's team met several challenging technical and business milestones including:

- ▼ Introduced proprietary NodeWatch™ Cluster Monitoring Tools
- ▼ Designed and manufactured our 1U custom dual Xeon chassis
- ▼ Hired four additional technical support staff and were awarded more contracts for on-site installations than ever before
- ▼ Added a senior scientist to our staff to further strengthen our core competency and enhance benchmarking capabilities
- ▼ Delivered several noteworthy clusters including 208 diskless nodes to University of Maine's Target Research and Development Center and 64 nodes to The Brain Imaging Research Center (a joint Center of Carnegie Mellon University and University of Pittsburgh); both clusters included Myrinet connectivity.

Products and services Microway will offer in 2003 include:

1) 64-bit solutions for HPC, based on Intel® Itanium® 2 and AMD Opteron™

In early January we shipped our first Itanium2 system. Currently we are accepting

orders for delivery of 1U dual Opteron development systems. Both platforms run 64-bit Linux and Microsoft® operating systems. Stay tuned for benchmarks and other information in future HPC Times editions.

2) Storage Solutions

Many HPC applications require multi-terabytes of data. Planning is also essential, as storage growth is often exponential. For example, many life sciences organizations are doubling their data every six months. Microway will expand our expertise and resources in the storage market specific to the needs of HPC, biotech and financial risk managers. Current relationships allow us to offer fully redundant, highly-available storage systems based on 2Gbit fiber channel technology. We are also evaluating storage directors with state-of-the-art capabilities such as access to any storage volume from any cluster node.

3) InfiniBand Connectivity

Significant advances have taken place in InfiniBand technology in 2002. Nearly a dozen hardware, software and solutions providers demonstrated InfiniBand-based products at SC2002. This is largely due to the availability of 4x InfiniBand switching and bridging silicon, which provides a compelling 10 Gb/sec, ultra low latency technology for cluster interconnects. The initial MPI performance data is impressive on small clusters (800 MB/sec sustained bandwidth with ten micro second latency). This year the “proof of concept” will require obtaining similar performance on very large clusters with the demonstrated ability to scale to clusters with thousands of nodes. Microway will continue to work closely with providers of various technology components to bring the right mix of hardware and software to the HPC community.

4) Integration and Consulting Services

Many markets including HPC, financial risk management and life sciences have dozens of commercial and open source software available to do various types of analytic work. In 2003 Microway will offer professional consulting services to our customers encompassing application expertise that extends beyond the design of fully integrated HPC clusters. These include recommendations for maximum benefit from

- ▼ Data warehousing
- ▼ Project management tools
- ▼ Network/data center design
- ▼ Bio-IT software custom installations for genomics, proteomics, chemical and drug discovery applications
- ▼ Lab integration and management systems

5) Integrated Solutions powered by Platform Computing products

Microway now offers workload management and grid computing software solutions to optimize computing resources. From our Clusterware-powered CW series to integration of LSF® 5 on any Microway cluster, all solutions are “grid-ready” for future system enhancements. All Microway customers will be able to maximize the performance of their clusters with these products. With the

inclusion of commercial applications like LSF 5 that “compete” with open source products like PBS, many users will need to consider the total cost of ownership to decide if the acquisition of commercial products from companies like Platform are more cost effective in the long run. See a related article in this issue of HPC Times, “Commercial vs. Open Source Application -- Tradeoffs”.

6) Windows and Linux Based MPI Professional Products

Microway will be integrating, MPI/Pro from MPI Software Technology. MPI/PRO is a commercial implementation of MPI and offers a significant improvement in performance over public domain versions of MPI. MPI/Pro is currently available for Linux and Windows® based clusters running Myrinet. It is expected that additional high bandwidth, low-latency interconnects will be supported soon. Watch this space for more information.

New Life Sciences Customers

Microway is pleased to announce the addition of two life sciences organizations to our customer base.

ALTANA Research Institute, the new technology-driven US research center of ALTANA Pharma, has purchased an 80-node dual Intel® Xeon™ cluster that will run BLAST for genomics and proteomics research. Microway's custom integration includes Platform LSF® to optimize the use of resources on the cluster.

Ge Zhang of Altana commented: “We selected Microway because of their reputation for integrating high quality hardware and their ability to deliver the LSF software at an attractive price.”

GlaxoSmithKline, a research-based pharmaceutical company, purchased 25 dual Xeon 2.8 GHz nodes. They will run in-house applications for numerical simulation and optimization of mathematical models for biological systems.

Valeriu Damian of GlaxoSmithKline commented: “It was a pleasure working with Microway. They offer competitive prices and a wide selection of configuration options. We wanted a dedicated 'computational appliance' i.e., a fast parallel machine with lots of memory and high speed interconnection for CPU bound applications with a competitive price. Microway was one of very few companies that quoted us the system we needed without having to pay a premium for features that we did not care about.”

**You Can
Count On It
By Stephen
Fried**

BPROC/Linux BIOS Research at LANL

One of the most interesting projects highlighted at SC 2002 was the development work being done at LANL to simplify the management of clusters. This is basically a throwback to the past. This article discusses the overall concept of BPROC and then discusses BIOS's in detail. It will be followed by another article on BPROC next month.

Alien File Servers

Today's Linux Beowulf paradigm traces its origins to the mid 80's, when scientists and

engineers at a number of universities and companies started experimenting with single board computers that they ran in parallel and linked to a PC. Eventually, the concept was commercialized by companies like Inmos and Thinking Machines. The Inmos Transputer needed I/O support to run applications. This was provided by the host system that frequently had a number of Transputer boards installed in it. The program Inmos ran to communicate with its Transputer farms was called an Alien File Server (AFS). The AFS stored the programs that were run on the Transputer farm along with any data that might be needed or generated. Communication was over Inmos Links, 2-bit serial lines that ran at 100 M-bits/sec almost 20 years ago! Each Transputer had four Links, which meant its maximum external bandwidth was 40 MB/Sec. The processors were 32-bit devices with built-in memory and link controllers which ran at 20 MHz and had roughly the same throughput as the best Intel processors then available. One nice characteristic of a machine hosted by an AFS, was that you didn't have to worry about each processor having a file system of its own, and all of the other issues that arise with a system composed of individual computers. To run an application you simply typed a command like RUNT8 followed by the name of your program and any command-like arguments to include, and the Transputer farm managed the rest. Another nice feature of such a system was that it booted in several seconds.

At SC2002, the folks at Sandia (Ron Minich and crew) demonstrated a miniature machine composed of several single board computers, which would boot and start up an application in several seconds. Their demonstration used a combination of tools, including their Linux BIOS and BPROC (Beowulf Distributed Process Space). BPROC basically is a throwback to the AFS of Inmos, while the Linux BIOS is a very fast booting BIOS. For those of you unfamiliar with BIOS's, I'll discuss this bit of code next.

BIOS stands for Basic I/O System; it is very basic. The first one was written by PC pioneer Gary Kildall in the late 70's for Intel 8080 development platforms. It was nothing more than another type of application called a monitor program. When your ancient Z80 or 8080 8-bit program crashed, it put you back into the monitor, from which you could probe memory to figure out what had happened. Kildall's BIOS was an afterthought. In 1975, he wrote a compiler for Intel called PLM-86, and as part of the project, he added a monitor, so you could run the code produced by the compiler. Intel customers liked the monitor code, and they included it with their early development systems. Eventually, Kildall discovered other uses for his monitor, and he marketed it as CP/M (which stands for Control Program and Monitor). The I/O routines in CP/M were extracted from it and built into a BIOS, so that it would be easy to port. Eventually IBM came along and bought a clone of CP/M called PC-DOS that was written by Seattle Computer and licensed to Microsoft. The BIOS came along as part of the deal. IBM published the sources to their BIOS as part of their new PC business model, which included making the IBM PC an open architecture machine. The BIOS of IBM was cloned by people like Compaq, Phoenix Technologies and AMI (to name a few).

Every time you turn on your computer, the first thing it does is run the BIOS initialization and test codes, which turn on the devices on your motherboard and initialize your system's memory and busses. (If you don't initialize memory at least once the parity and ECC systems will not be able to function). At that point in time, if you run MS-DOS, your machine will take advantage of the BIOS routines for any I/O done. If you run Linux or NT, the BIOS code will now be used to read a bootstrap loader into memory and then jump to it. Once the bootstrap loader is up and running, it has all of the I/O resources it needs to do things like load the image of the Linux Kernel. Next, your machine basically wipes clean the image of the BIOS, and starts to run using the device driver routines built into Linux. You might ask, why doesn't Linux just use the BIOS routines? It turns out that at the time IBM wrote its BIOS, the peripheral chips used in the PC were buggy and IBM was forced to avoid using interrupts.

Without interrupts it's difficult to manage more than one task at a time. This meant that the original PC's were single threaded, which was also the case for the BIOS and MS-DOS. Linux is a multi-tasking OS, in which the BIOS gets replaced by the kernel and its device drivers.

The bottom line is that you don't need a BIOS in a PC, except for the first few seconds, and if you replace it with code that you have the sources for, you can do all sorts of interesting things. What Ron Minich of Sandia did was to use his BIOS (which is basically a few lines of assembly language along with routines extracted from the Linux kernel) to make it possible to boot a PC in a few seconds, instead of the minutes that it typically takes. With this accomplished, he was now ready to run his computer farm, using techniques similar to those employed by Inmos that made it possible to boot a Transputer farm in a few seconds.

Next time we'll talk about BPROC, the code Sandia used to replace the Inmos AFS.

**Tech Notes
By Nina
Nitroy**

Commercial Compiler Options for Clusters

For the best performance and optimization, one can choose from several commercial compilers available for use on clusters. See the following URLs for specific features. Optimization features you should look for include support for Intel 64-bit processors or AMD Opteron 64-bit processors. If you have no specific preference, Microway will configure your cluster for use with the GNU C and Fortran compilers upon request. Among the features available within most commercial compilers are OpenMP support (parallel threads), debuggers, Fortran 95 support, vectorization, GNU C library compatibility, and processor-specific optimizations (such as MMX, SSE, SSE2, 3DNow). Many have cluster packaging available for additional tools related to Beowulf clusters. Pricing varies.

Intel

Fortran and C++ compilers

developer.intel.com/software/products/compilers/

Current Version 7.0

Includes Itanium2 64-bit optimizations

Portland Group Inc.
PGI Fortran and C++ compilers
www.pgroup.com/
Current Version 4.0
Includes AMD Opteron 64-bit optimizations

Absoft
Fortran and C++ compilers
www.absoft.com/index2.html
www.absoft.com/workstation.html
Current Version 8.0
IMSL math libraries available for purchase

Lahey Computer Systems, Inc.
Lahey/Fujitsu Fortran 95
www.lahey.com/
Current Version 6.1
Includes MATLAB compatibility

Compaq
(For Alpha systems and clusters)
Compaq Professional LINUX Developer's Software
Fortran, C++, debuggers, Compaq math libraries, visual threads
www.support.compaq.com/alpha-tools/index.html
Current Version 7.2

Customizing Your Microway Cluster – How It Works

Microway configures Beowulf clusters with a variety of software upon request:

Message passing:

MPI (MPICH, LAM MPI)
PVM

Batch Queuing, Scheduling, Load Balancing:

OpenPBS (open source) www-unix.mcs.anl.gov/openpbs/
PBS PRO (commercially supported) www.pbspro.com
Platform Computing software products www.platform.com/products/
MOSIX www.mosix.cs.huji.ac.il/

Cluster Management and Administration:

Microway Cluster Management Software MCMS (installed on all clusters)
Platform Computing Clusterware www.platform.com/products/Clusterware/index.asp

Parallel Virtual File System (PVFS): parlweb.parl.clemson.edu/pvfs/

Network Teaming

Channel Bonding: www.beowulf.org/software/bonding.html

Software RAID

Basic Firewall
OS Services (NFS, NIS, Apache, proftpd, mail server, etc.)
LINUX (RedHat, Debian, Mandrake, and others)
Custom, up-to-date kernels
Choice of file systems and custom partitioning
Scyld Beowulf
Network Boot Diskless Nodes
and more...

Upon receipt of a cluster order, Microway provides you with a "Cluster Questionnaire." Completing the questionnaire provides Microway with information about your specific needs and requests for the software build.

**Parallel
Thoughts
By Jay Owen**

Commercial vs. Open Source Applications – Tradeoffs

In December Microway signed an agreement expanding our relationship with Platform Computing to include reselling all Platform products including their flagship software, LSF® 5. As we add partners to our offering, users want to understand why someone would purchase commercial software when an open source alternative is readily available. In several instances, a broad look shows that commercial products are more cost effective than their "free" counterparts. Here is a brief comparison of Open PBS vs. LSF 5.

The argument for Open PBS centers largely around the acquisition cost...it is zero. Open PBS can be customized by end users familiar with scripting languages. If the user does not have this expertise in-house, experienced developers and administrators are available in the market. However, once the user contracts services, then costs have been incurred. Even if in-house expertise exists there is a "soft cost" associated with their time. There are other considerations relating to growth and scale that might make the Open PBS route more expensive.

Another issue exists with ongoing maintenance. Without the proper resources on-site, configuration and maintenance are difficult to accomplish. In critical cluster applications, predictable resolution of problems is essential. Companies like Platform stake their existence on providing reliable products with predictable, effective support. Support occurs in the public domain unpredictably. Inordinately long "down time" relates to higher opportunity costs or the lost value associated with a scientist unable to work.

Platform LSF 5 provides scalability at the single cluster level. Platform also offers supported add-in products for special user needs. Just a few of them provide features such as:

- ▼ Real time resource management for parallel MPI jobs on a single cluster (Parallel)
- ▼ Industry specific modules for industries such as financial services (Symphony)
- ▼ Performance management, resource management and license management tools

LSF Multicluster or the Platform Globus Toolkits are the Platform products that provide scaling from a managed single cluster to a grid environment. These point offerings are all supported by a single vendor, again insuring that a migration is predictable from cost, technology and timing perspectives. If the trend at the installation is to build larger clusters

or a pervasive computing environment that spans multiple clusters, then using Open PBS at the outset may complicate future growth or administrative requirements.

Some features uniquely available in Platform LSF are job prioritization, the ability to add fault tolerance and recovery capabilities to the cluster. Numerous mechanical design and life sciences applications have hooks that allow integration into a cluster with LSF.

If the user's anticipated needs are fairly stable, experienced technical personnel is available at little or no cost (opportunity cost or actual out-of-pocket expenses) and the desired functionality is basic/small scale then an Open PBS solution may make sense. However, if the user anticipates upgrading the system and needs to minimize the need for technical expertise, purchasing commercial software can result in the lowest total cost of ownership solution compared to the lifetime incremental cost of additional service and support.

If you have a question about Platform's products and how they might apply to your computing environment, please contact a Microway representative for a consultation with experts from Platform and Microway...

**HPC
Forum**

What do you predict will be the most promising new technology in the HPC industry in 2003?

Share your thoughts with us by emailing hpc@microway.com and we'll post some of the answers in next month's newsletter. Creating a community of questions and answers.

To view Microway's HPC News online or receive it via email www.microway.com