Installation of a New Linux Cluster Facility for Massive Parallel Computing

M. Schubert, F. Grüner, and W. Assmann

A new Linux cluster has been set up for massive parallel calculations in the field of atomic interactions and dosimetry of beta radiation. The main goal of this configuration was to get the most calculation power out of the given money.

1. Hardware

The hardware and the initial software installation was bought from a company specialized on building clusters [1]. The cluster has 26 AMD Athlon MP 2000+ with 1.66 MHz. These CPU's are combined on dual processor boards in 13 standard midi towers. Each node is equipped with 1 GB DRAM memory and two Ethernet controllers. Only the front end is equipped with a third network adapter and two 120 GB hard disks. To save money for more calculation power, existing racks and infrastructure have been used to house the nodes (see figure 1).

Fig. 1: The two racks containing the 12 nodes, the front end and the two network switches of the Linux cluster.

Figure 2 explains the schematics of the cluster. The two Ethernet adapters per node are used to separate the calculation traffic (MPI etc.) from the bulk traffic (nfs, ssh, etc.). For both networks two separate store and forward switches are used. This increases the network throughput and reduces the network latency. This is a cheap alternative for low latency networks like Myrinet, SCI and Infiniband, which are very expensive (roughly 40% of the cluster costs) and only needed for highly interacting codes, where the latency of the network is a critical parameter. The codes running on this cluster are loosely coupled.

The 12 nodes are build diskless and booted by PXE [5] and nfs-root [6]. This reduces the installation and administration overhead. The front end has two IDE disks, which are mirrored in Raid 1 mode for data safety.

To protect the cluster against any unwanted network activities, the nodes are only reachable over the front end, which has the function of a gateway and a firewall using iptables.

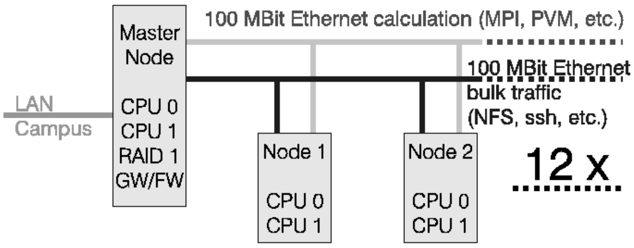

Fig. 2: The schematic overview of the configuration and interconnections of the cluster.

2. Software

The operating system of the cluster is RedHat Linux 7.3. The installed compilers are GNU C++ and Intel C++ Compiler. For the message passing the MPI [3] standard is used in the MPICH [4] implementation. As user code a Geant4 based parallel implementation and several other codes are used.

3. Performance

For a performance check the High Performance Computing Linpack Benchmark (HPL) [7] was used. It resulted in a peak performance of 18 GFlops/s. The hardware and software stability of the cluster was excellent till now. The downtime due to administration purposes was around half a day. The was no unplanned shutdown caused by hardware or software defects. The average idle time till now is below 0.7%. This substantiates, that this cluster is currently operated at full capacity by its current users.

4. Outlook

For further calculation projects from different users an extension of the cluster was already ordered. This will double the calculation power from April 2003 on.

References

[1] MEGWARE Computer GmbH, http://www.megware.de/
[2] Geant4, http://geant4.web.cern.ch
[3] The Message Passing Interface (MPI), http://www-unix.mcs.anl.gov/mpi/
[4] MPICH-A Portable Implementation of MPI, http://www-unix.mcs.anl.gov/mpi/mpich/
[5] Pre-Execution Environment, http://syslinux.zytor.com/pxe.php
[6] Root over NFS, http://www.tldp.org/HOWTO/Diskless-root-NFS-other-HOWTO.html
[7] High Performance Computing Linpack Benchmark, http://www.netlib.org/benchmark/hpl