A DISKLESS SOLUTION FOR LCG MIDDLEWARE

F. Acernese[1,2], F. Barone[2,3], R. De Rosa[1,2], A. Eleuteri[1,2], R. Esposito[2], L. Giordano[2,4], P. Mastroserio[2], L. Milano[1,2], S.Pardi[2,4], G. Russo[1], F.M. Taurino[5], G. Tortone[2],

[1]*Università di Napoli "Federico II", Dipartimento di Scienze Fisiche, 80126 Napoli, Italy*

[2]*INFN – Sezione di Napoli, 80126 Napoli, Italy*

[3]*Università di Salerno, Dipartimento di Scienze Farmaceutiche, 84084 Salerno, Italy*

[4]*Università di Napoli "Federico II", Dipartimento di Matematica ed Applicazioni "R.Caccioppoli", 80126 Napoli, Italy*

[5]*INFM – Unità di Napoli, 80126 Napoli, Italy*

Abstract

The INFN-GRID project allows experimenting and testing many different and innovative solutions in the GRID environment. In this research ad development it is important to find the most useful solutions for simplified the managment and access to the resources. In the VIRGO laboratory in Napoli we have tested a non standard implementation based on LCG 2.6.0 by using a diskless solution in order to simplify the site administration. We have also used a shared file system, thus simplifying the submission of MPI jobs to the grid. Finally we tested the use of a custom kernel with the OpenMosix patch in order to enable the dynamic load balancing.

INTRODUCTION

The detection of gravitational waves (GW) is one of the most interesting fields of the modern physics: it will provide a strong proof of the general relativity theory, opening up a completely new channel of information on the dynamics and evolution of astrophysical objects [1].

Within this framework, the large-scale terrestrial interferometric detectors like VIRGO [2], LIGO [3], GEO [4] and TAMA [5], will have a prominent role. In fact, these detectors will operate with large detection bands, typically spanning from 10 Hz up to 10 kHz, with a sensitivity of about 10^{-21} h/(Hz)$^{1/2}$ at 100 Hz, where h is the gravitational strain. In addition the very long baseline space interferometer LISA [6] will explore the detection frequency band from 10^{-5} up to 1 Hz.

For all these detectors, but especially for earth-based antennas, the main problem to solve in data analysis is the expected low signal-to-noise ratio (SNR). The low value of this quantity is due to the intrinsic weakness of gravitational signals with respect to the instrumental noise.

To overcome this problem much work is ongoing in the development of suitable data analysis techniques. When the expected signals shape is known, the most promising technique seems to be the matched filtering, i.e. the correlation of the detector output with a set of theoretical waveform templates.

In the case of the VIRGO antenna the required computing power for detecting gravitational waves generated by a coalescing binary systems is about 300 GFlops for masses ranging from 1.4 to 50 solar masses, assuming a signal-to-noise recovery of 90% [7][8]. This reach can be obtained using computer farms composed by several nodes connected to each other through the network. This technical solution represents the only current possible way for an on-line data analysis. Of course a more accurate analysis can be performed off-line. In this case there are no time constraints and therefore the number of used templates can be easily increased and other parameters can be also included in the model. For this reason the off-line phase usually requires a very large amount of computing power that can be obtained only by adding more computing resources. The direct consequence of this approach is that a problem of optimum algorithms development must take into account the farm architecture and configuration. For this reason, we implemented a very versatile and modular computing power tool in Napoli.

In order to simplify the administration of the laboratory beowulf cluster, the Naples group has implemented a multi-boot configuration with a centralized web interface.

Every node can boot from network and work as a diskless node of the local cluster, otherwise local boot is enabled in so that the node is a Worker Node of the GRID site INFN-NAPOLI-VIRGO.

Based on this experience, we have tested a diskless solution for the LCG software in a personalized configuration, in order to simplify the upgrades of the middleware and introduce the openMosix patch in GRID

HARDWARE ARCHITECTURE

The Beowulf cluster is a heterogeneous farm composed from 20 nodes that can be divided into two main homogeneous hardware subsets. Furthermore another 5 machines are used as basic elements for the GRID infrastructure management and for geographically distributed computing; this will be explained later in the paper. The two subset are:

Super Micro 6010H subset: 12 nodes each equipped with two Intel Pentium III 1 GHz, 512 MByte RAM, a local 18 GByte SCSI disk and 2 integrated network connections on boards (Fast Ethernet 10/100). The first node is equipped with a further Giga Ethernet board, as it may act as the farm master.

APPRO 2114Xi subset: 8 nodes each equipped with two Intel Xeon 2.4 GHz processors, 1024 MByte RAM, a local 60 GByte IDE hard disk, mother board TYAN, and two integrated network connections on board (Fast Ethernet and Giga Ethernet).

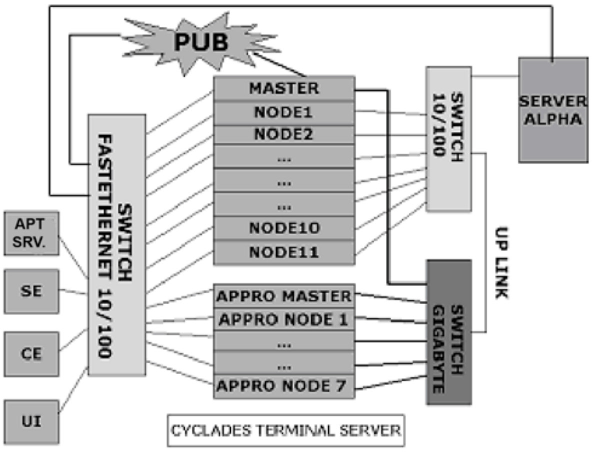

Fig.1
Hardware architecture of the Farm

The farm has two independent networks: the first one is used to optimize the data transfer speed among the farm nodes, is split into two subsections, consisting of a Fast Ethernet switch and a Giga Ethernet switch. An uplink connects these two switches to implement an equivalent single private network. The two private network subsections are necessary because the Super Micro nodes have Fast Ethernet links, while the APPRO ones have Giga Ethernet links.

The other network is used for public access to the farm and for its management. It consists of a Fast Ethernet switch to which all the farm nodes are linked, together with an Alpha Server and all the basic GRID elements (the APT server, the Storage Element, the Computing Element and the User Interface). A Terminal Server (Cyclades S2000) is also connected to the public network for the farm remote management, while all the nodes are also connected to the Terminal Server through serial links. A scheme of the Farm Architecture is shown in Figure 1.

SOFTWARE ARCHITECTURE

For the test we have used only 3 APPRO node of the farm. The chosen operating system is Scientific Linux 3.0.4 with OpenMosix kernel 2.4.20. The middleware used is INFN-GRID 2.6.0.

The diskless architecture is obtained by installing a worker node with the typical procedure and standard parameters provided by installation guide of INFN-GRID. This machine play the role of MASTER node in the diskless configuration.

After the installation we performed our personal implementation by using a different kernel with OpenMosix patch.

The MASTER node share the root with the other member of cluster by using ClusterNFS that provide a set of patches for the "Universal NFS Daemon" (UNFSD) server, to allow multiple diskless clients to NFS mount the same root filesystem by providing "interpreted" file names.

When a slave node goes up, at the boot stage it asks for the kernel through TFTP from the node MASTER; the kernel is installed mounting the *root* directory and all the packages via NFS. The local disk of the node is used as a swap area.

A large data transfer with OpenMosix or MPI on the network may cause the failure of some nodes. The solution is to have two networks, the first one to mount the *root*, the disks, to access the machines from outside, etc., and a private network for the parallel computing data transfer. In this way the farm becomes very stable and performant.

RESULTS

The first result obtained by this kind of configuration is the introduction of the concept of SSI (Single System Image) in GRID Environment, that is a cluster of workstation that seem an unique system, that support the classic parallel job with static allocation and the dynamic load balancing with OpenMosix.

The main goal achieved is the simplification of the administration of the site. With this configuration we need only upgrade a machine for update all the worker node of the site.

Finally we explore also the possibility to send a MPI job in a simplest way.

We say that, in order to execute a MPI job on a LCG farm configured in a standard manner, we have to find the worker node available on the farm, creates the working directories on all the nodes allocated for parallel execution and copies the needed files on all the nodes allocated for parallel execution. After this operation we can execute our job.

With our configuration, we taking advantage from the shared home of the Worker Node; so we can execute a MPI job without copying the executable on the different Worker Nodes involved in the computation, ande the classical script for a parallel job it becomes as simple as:

```
#!/bin/bash
  chmod 755 `pwd`/mpi_job
  mpirun -np 7  `pwd`/mpi_job
  Conclusion and future works
```

CONCLUSION AND FUTURE WORS

The test on the diskless configuration of LCG has shown the possibility to obtain numerous advantages in the administration and the use of the grid resources. In the next few months we propose to extend the configuration to all the farm nodes and to execute a testbed on the INFN-GRID in order to test the system stability.

REFERENCES

[1] C. W. Misner, K. S. Thorne, J. A. Wheeler, Gravitation (Freeman & Co., San Francisco, 1973).

[2] C. Bradaschia et al., "The VIRGO Project, Final Design of the Italian-French large base interferometric antenna of gravitational wave detection", Proposal to INFN Italy and CNRS France, 1989, 1992, 1995.

[3] R.E. Vogt, R.W. Drever, F.J. Raab, K.S. Thorne, "Proposal for the construction of a large interferometric detector of gravitational waves", Proposal to the National Science Foundation, California Institute of Technology, Pasadena, California, USA, 1989.

[4] Hough et al., "Proposal for a joint german-british interferometric gravitational wave detector", MPQ 147, Max Planck Institut für Quantenoptik, Munich, Germany, 1989

[5] unpublished, see http://tamago.mtk.nao.ac.jp (1996)

[6] P. Bender, et al. "LISA: Laser Interferometer Space Antenna for the detection and the observation of gravitational waves},MPQ 208, Max Planck Institut für Quantenoptik, Munich, Germany, 1996.

[7] B. Owen, 1996, Phys. Rev. D, 53, 6749.

[8] P. Canitrot, L. Milano, A, Vicere', VIR-NOT-PIS-1390-149, 2000

[9] I. Foster, 2002, Physics Today, 55, 42

[10] L. Blanchet, T. Damour, B. R. Iyer, 1995, Phys Rev. D, 51, 5360

[11] C. W. Helstrom, Statistical Theory of Signal Detection, 2nd ed. (Pergamon Press, London, England, 1968)

[12] Chervenak, I. Foster, C. Kesselman, C. Salisbury, S. Tuecke, 2001, Journal of Network and Computer Applications, 23, 187