The LAFEX Risc Based Farm Solution for CPU Intensive Applications in HEP

G. A. Alves [a,1], M. Mendes [a], M. Miranda [a], A. Santoro [a,b], C. Silva [a], M. Vaz [a]

[a] *LAFEX/CBPF*
R. Dr. Xavier Sigaud, 150
Rio de Janeiro, RJ, Brazil

[b] *Fermilab*
P.O.Box 500, Batavia, Il
60510, USA

The problem of increasing CPU demand of High Energy Physics applications have been dealt by setting up parallel processing "farms", in which the processors work in cooperative mode. Commercial solutions to this problem usually requires high investiments in purchase and maintenance. We have built at our home institution a very affordable system, using off-the-shelf components, which besides of fulfilling our computing needs, allowed us to give a significant contribution to the Monte Carlo production effort for the Fermilab DØ experiment.

Key words: High Energy Physics; Parallel Processing; Computing Farms

1 Introduction

High Energy Physics is well known for its event oriented and CPU intensive type of applications. Event Reconstruction of a large hadron collider detector, like the DØ detector at Fermilab, usually takes around 1000 Mips/second for a single event. The situation is even worse for the case of Event Simulation using the Monte Carlo Method, including the Simulation of the Detector Response, which in this kind of environment can have more than a factor of 10 over the reconstruction case. It is clear from the current scenario, that the CPU needs for the future collider experiments will be much bigger than the present ones. On the other hand, the event oriented character of HEP leads naturally to the parallel processing paradigma. Available commercial solutions for massive parallel processing, like the Challenge and SP2 machines, are currently not very attractive in terms of cost/performance, and its high costs for maintenance can only be afforded by big institutions, reinventing the mainframe. We have proposed, and built at our Institute(LAFEX/CBPF), a solution for the high CPU demand of the High Energy Physics groups at LAFEX, using a FARM architecture composed of standard off-the-shelf components.

2 Farm Architecture

A schematic diagram, showing the farm architecture can be seen in Fig. 1. The Farm consists basically of 30 RISC based slave nodes(IBM-7248 Model 43P), divided into 3 farmlets, each one attached to a disk server(IBM-7009 Model C20) via a ethernet switch, which filters the packet traffic so that the system can be configured for running as 3 independent farms. The servers have a high disk storage capability(20 Gbytes/server), plus magnetic tape storage capability, using high density 8mm tape library systems(IBM-7331 Model 205). All slave nodes are connected, via serial port, to a console server(IBM-7011 Model 25T), which besides of displaying errors and console messages, performs the functions of monitoring errors, load of the system software, software installation, etc.

[1] gilvan@lafex.cbpf.br

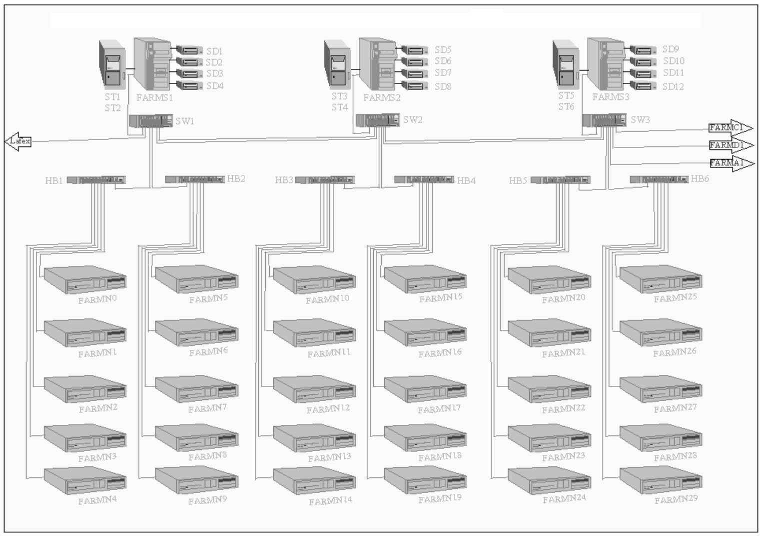

Fig. 1. The LAFEX/CBPF Farm layout. Every group of 10 nodes are connected to a File Server via a Ethernet Switch, making a farmlet configuration module. Modules can be added to increase the power of the system.

The system was conceived to be highly modular, in the sense that additional farmlets can be added, to increase its processing power, without affecting the base architecture.

Coupled to this system is a Client/Server software interface[1] for event oriented parallel processing. Using this software and the farm's modularity, allows one to configure "virtual farms", that is to say, farms composed of some processors of the LAFEX farm together with other processors elsewhere in the world.

This combination of hardware and software represents one possible solution for the high demand of Monte Carlo event simulation of the DØ Collaboration. Besides of doing Monte Carlo production, part of the system can be dynamically allocated to run user applications, resuming the production job afterwards.

Another aspect of this system is that it is very affordable for a ordinary institution, without the extra costs that come with all of the specialized systems, e.g. special buses, switches, etc. Even using only standard components we get a cost/performance rate which is at least a factor of 2 better than the proprietary commercial solutions.

3 Conclusions

We built a modular farm using just standard components, which can be expanded according to the system demand. This solution can certainly be used for other fields with computing requirements similar to High Energy Physics event simulation and reconstruction. For applications with this characteristics there is no need for a tightly coupled system, and the processing nodes can be spread over the world in a virtual farm without affecting the global performance of the system.

References

[1] M. Miranda *et al.*, *WWW Oriented Remote Job Submission, Monitoring and Management over Internet* - This conference proceedings