

# Meteorological Applications Cluster using OSCAR 2.3.1

**Richard Miguel San Martin, SENAMHI**  
**rmiguel@senamhi.gob.pe**

## Abstract

OSCAR is an application that makes possible build a cluster easily. Is used when we have a great number of nodes and the task of build a cluster installing node to node is so hard. OSCAR (Open Source Cluster Application Resource) is a project managed by Open Cluster Group which is a group of people that works for OSCAR is very popular.

This work is about how is possible build a cluster using OSCAR with low investment, and limited resources in developing countries.

This cluster was build for meteorological applications on a project of climate change research, however use another scientific applications is possible without problem.

## Introduction

CPN (Numerical Prediction Center) is an office belongs to SENAMHI [1]. This office is in charge of to run numerical models for prediction of weather and climate and we have some models as ETA model RAMS model [2] each one applied to different regions and with different initials conditions too.

SENAMHI is involved in a project for to make climate scenarios with forecasts to long time. The time proposal is to 50 years, it means what we require a great calculus capacity.

The computation capacity of SENAMHI is limited because we have two workstations Alpha with 4 processors each one only and now these workstations are working in your maximum capacity running other models and is producing forecasts daily for our users.

After evaluate the options available, we take the decision of build a cluster based on Linux and Intel processors for to realize this task on our project.

This work is about our experiences in the evaluation of options and how we take the decision of implement a cluster Beowulf using OSCAR for the installation.

## 1. Design of Cluster.-

The design of this cluster is based in the clusters Beowulf methodology with adds about performance and improvement of network that will be detailed in this work.

### 1.1 Purpose.-

This cluster was designed as result of an evaluation of different proposals, since the implementation of a cluster using Alpha processors to commercial solutions as Blade server of HP or Sun. The intention was to build a robust infrastructure to be able to use applications of high performance with low cost, that is to say, to obtain good performance with limited resources. Finally we take the decision of to realize a solution based in Intel processors using Linux as operative system. To continuation some reasons on which we based our decision.

### 1.2 Objectives.-

- To obtain a technological solution that allows us to improve the current capacity of calculation of Numerical Prediction Center and that simultaneously is scalable, secure, powerful and reliable.
- To realize forecasts of weather and climate in the long term obtaining minor times of response to the reached ones at present with the installed infrastructure.

### 1.3 Alternatives of Solution.-

- a. Update of Equipment.
  - Acquisition of more discs to optimize the capacity of storage.
  - Update of operating systems of Tru64 v.4.0f to 5.1.
  - Update of processors.
  - Increase of memory.

- b. Acquisition of new equipment.
  - To continue using technology RISC with equipments of cluster HP - Alpha.
  - To acquire equipments RISC of other marks as SUN or IBM.
  - Acquisition of massive storage system.
  - Probable migration of applications to adapt itself to new operating systems.
  - Training in new technology.
- c. Implementation of cluster Linux based on Intel technology.
  - Group of equipments based on processors Intel.
  - Use of operative system Linux and applications of free distribution.
  - Permanent infrastructure.
  - Capacity of calculation equivalent to equipments of other technologies.

#### 1.4 Why a cluster Linux?

- Because Linux is a stable and secure operative system.
- Because a cluster using Intel processors and Linux is very much cheap that another commercial solution as HP, IBM, Sun.
- The use of software of free distribution allows to build a stable, administrable solution and of low cost. Besides the use of free software would allow us to reach level of experience and knowledge of the technology very higher than other solutions.
- Linux allows us to obtain a reliable, robust solution that allows us to reach comparable performance to another commercial solutions but to a much smaller cost.

#### 1.5 Proposed solution.-

Our solution is based in Linux as operative system and Intel processors as technological platform:

- 30 processors Intel Pentium IV 2.4 GHz.
- 256 Mb RAM (8 GB Total)
- 60 GB HD (1.8 TB total)
- Network card 10/100 and 10/100/1000.
- 2 Switch 3Com 48 ports.
- Red Hat Linux.
- Mpich passing of messages library.

- RAMS model.
- Fortran PGI compiler.

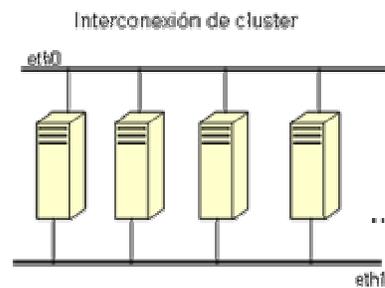


Fig. 1. - Cluster interconnection

Our cluster is composed by 30 nodes connected in two networks. One network (10.0.0.255) is for IPC and process communication. The network (10.0.1.255) is for transmission of data on the network (I/O).

In addition to it exist two server nodes or master node those who take charge sending the processes to other nodes components of the cluster.

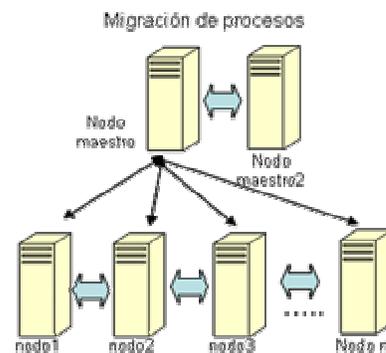


Fig. 2. - Process Migration

Our cluster use mpich for process communication, our application is a weather and climate model that uses Fortran PGI for your compilation.

The process communication between nodes is using mpich and the communication of data is using NFS and mounting discs in accordance with the need of space needed by our application.



Fig.3.- Rack with 19 nodes

## 2. Using OSCAR.-

Our principal problem for the installation of the cluster was how to realize it in 30 nodes without using a lot of time in it and also simplifying the tasks of configuration of the applications necessary to construct it.

Initially we did realize tests using other tools as SCE [3] and ROCKS [4], which were evaluated in terms of facility of installation, configuration and additional tools.

After having realized these evaluations, we decide to use OSCAR for the following:

- OSCAR allows us to have an interface that shows us step by step the sequence of the installation of the tool.
- It allows selecting the packets necessary for the installation in accordance with the need of the user.
- Configure automatically packets as mpich, lamp.
- It is very easy to add more nodes later.
- It is possible to have multiple configurations of clusters in the same server.

### 2.1 Installation.-

Before realizing the installation of OSCAR, the following steps are realized:

- Installation of Red Hat 8.0 standard version in server machine.
- Create of directory in */ftpboot/rpm*, here should be the packages of Red Hat (rpm).
- Consider the requirements of space on disk.

- Download and unpack the OSCAR package on the server.

After realizing the steps of configuration and compilation detailed in the Installation Guide of the documentation, we realize the steps of configuration of OSCAR for the nodes clients.

Initially we were using ssh to realize the connection between the nodes, but due to the need of our application that, on having used rsh, it was improving notably the time of response, realizing a change in */opt/mpich-1.2.5.10-ch\_p4-gcc/bin/mpirun*. We have changed the line `RSHCOMMAND="ssh"` by `RSHCOMMAND="rsh"`.

The content of our *sample.disk.ide* file is:

```
/dev/hda1    100    ext3    /boot defaults
/dev/hda5    512    swap
/dev/hda6    *      ext3    /      defaults
nfs_oscar:/home --    nfs    /home rw
```

Additionally we have installed Fortran PGI as compiler for our application.

### 2.2 Other considerations.-

- We choose build our cluster using Red Hat Linux v. 8.0 because we think that we have certain experience managing this distribution and because OSCAR guarantees full support using this distribution.
- During the installation we had problems neither with respect to the hardware nor problems derived from OSCAR.
- The modules of the network cards were recognized without problem and in some cases when it did not happen this completed the installation manually loading the necessary modules.
- The installation of the nodes was realized one to one using a disk with the image. For this we use directions ip static. The network cards that we use are 3Com 3C509BTX and some machines have 3Com 3C996-SX Gigabit Ethernet.
- In resume, we have a cluster with 30 nodes in a network fast Ethernet with possibility of migrating to gigabit changing the switch and some of the current cards.

### 3. Our application.-

Our application is a numerical model that being based on differentials equations and associating it with different variables realizes an analysis of the state of the atmosphere and allows us to infer forecasts in the long term. The model RAMS was developed by ATMET [5] with the purpose of realizing forecasts in domains with very high resolution.

The development of a parallel version of RAMS began in 1990 and has continued since then. While RAMS contains a number of features which makes it amenable for modification to distributed memory platforms, such as the use of explicit numerical schemes in the horizontal finite-differencing algorithms, there are also a number of features that make the parallelization more difficult. These include a large number of numerical and physical options. The major complicating factor, however, is the interactive (or two-way) nesting procedure that allows the user to specify any number of telescoping grids or even moving grids that could float through a larger grid while calculating, for example, transport and dispersion of a pollutant or the propagation of a thunderstorm. This also allows a high resolution forecast for a target area while simultaneously providing coarser resolution forecasts for a much larger surrounding area.

The code was first developed using PVM in a master-node configuration where all nodes only communicated with the master process. As communication hardware became more sophisticated, this was seen as a bottleneck. The code has subsequently been changed to node-to-node communications using MPI, retaining the concept of the master process for output files and dynamic domain decomposition.

For parallel execution serial code was divided in two types:

Master process: (initialization and all input/ output functions).

Node compute processes (all computation).

Is not necessary that the files output are stored in the nodes since these come back to the main node.

This makes possible to keep storage centralized, which facilitates the backups.

Our task is to realize forecast of climatic change using the out of this model for next 50 years. This task involves normally a demand of calculation that with the previous infrastructure that we had had been reached in major time.

### 3.1 Some Results.-

Dominio Piura – Vientos (m/s) y Temperatura

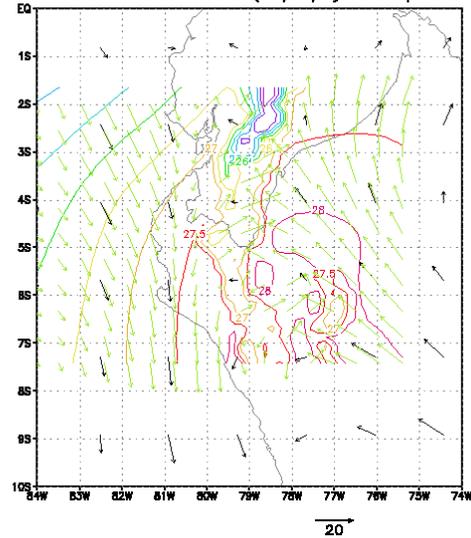


Fig. 4. Forecast of winds and temperature on Piura domain for 2016 April.

The project involve the study of two domains or regions from Peru, Piura and Mantaro domains. Piura and Mantaro are two regions from Peru very much affected by the phenomenon El Niño and therefore they are a center of attention of scientists and meteorologists for this type of studies.

As there was said previously, the target is to strengthen capacities of tasks of prevention before natural disasters for what we need to realize forecasts in the long term on the above mentioned regions.

Dominio Mantaro – Vientos (m/s) y Temperatura (C)

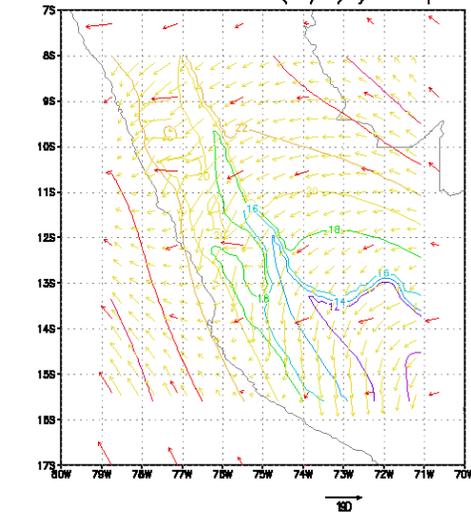


Fig. 5. Forecast of winds and temperature on Mantaro domain for 2016 April.

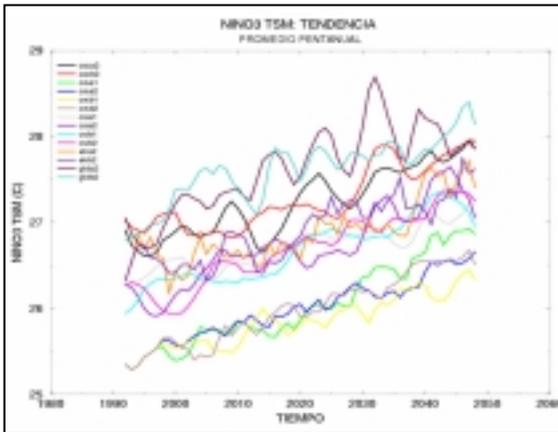


Fig.6. Tendancy of temperatures of sea during phenomenon of El Niño (1992 - 2050)

#### 4. Performance of our application on our cluster.

In the first tests of performance with our application we configure the RAMS model to realize a simulation of forecast of 1 day for the domain Peru to 60 Km and the result can be appreciated in the fig 7.

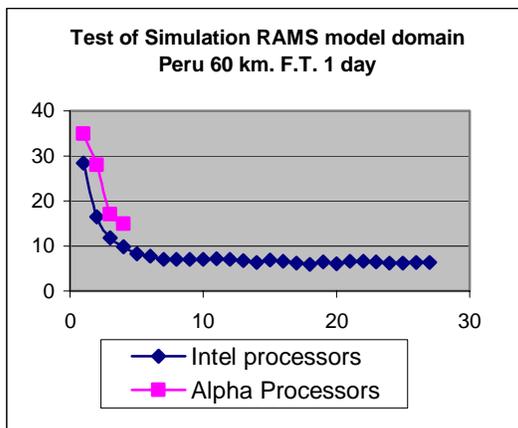


Fig. 7. Test of simulation Forecast Time: 1 day.

The same configuration of domain is used with the same time of forecast and a simulation is realized using 1 Alpha processor and 1 processor Intel later, obtaining a time of response of 35 minutes in the first case and of 28 minutes in the second one.

Later we were adding nodes to the cluster and repeated the experiment up to reaching 27 nodes in this test and we have achieved as less time of

response appears in the figure using 16 nodes and with a time of 5 minutes and a half.

After the seventh or eighth node the time of response does not diminish significantly therefore, we decide to realize a logical division of the cluster in groups of 7 machines executing different process for different times in each of the groups of machines in parallel form. This solution was the most ideal for us but not necessarily it might be the best for another type of applications.

#### 5. Future tasks

- To assure high availability in the master node finishing this way the vulnerability in the single point of failure.
- To optimize the performance of the cluster across tasks of tuning of the operative system.
- In the short term to migrate of a network fast Ethernet to gigabit Ethernet.
- To realize other tests of performance with other applications of scientific use to determine the potential of this cluster in other areas of the science.

#### 6. Conclusions

- Implementing clusters based on Linux is a viable alternative for developing countries.
- OSCAR facilitates very much the work of installation and configuration of the cluster.
- To remember that in this version of OSCAR is necessary to use of preference hardware with similar characteristics.
- OSCAR is a flexible tool, easy to use and with every time major followers.
- The obtained results indicate that not always there is obtained a decrease of the time of response proportional to the number of processors used.
- The costs of implementation are very low in relation to other solutions.

## 7. References.-

- [1] SENAMHI Servicio Nacional de Meteorología e Hidrología - National Service of Meteorology and Hydrology.
  - [2] RAMS Model  
<http://rams.atmos.colostate.edu/>
  - [3] SCE (Scalable Cluster Environment)  
<http://sourceforge.net/projects/sce/>
  - [4] ROCKS  
<http://www.rocksclusters.org/Rocks/>
  - [5] ATMET  
<http://atmet.com/>
- Installation Guide OSCAR 2.3.1
  - OSCAR Cluster User's Guide.
  - How to Build a Beowulf: A Guide to the Implementation and Application of PC Clusters (Scientific and Engineering Computation)
  - Craig Tremback and Robert Walko, "RAMS in Parallel Computing." (2000)