Grid Computing for Mesoscale Climatology: Experimental Comparison of Three Platforms

H.F. de Campos Velho[1], A.J. Preto[1], S. Stephany[1], E.R. Rodrigues[1], J. Panetta[2], E.S. Almeida[2], R.P. Souto[2], P.O.A. Navaux[3], T.A. Diverio[3], N. Maillard[3], and P.L. Silva Dias[4]

[1] Laboratório de Computação e Matemática Aplicada – LAC/INPE
[2] Centro de Previsão de Tempo e Estudos Climáticos – CPTEC/INPE
INPE – Instituto Nacional de Pesquisas Espaciais, S. J. Campos, Brazil
[3] Instituto de Informática – II/UFRGS
UFRGS – Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[4] Departamento de Ciências Atmosféricas – IAG/USP
Instituto Astronômico e Geofísico (IAG)
USP – Universidade de São Paulo, São Paulo, Brazil

Abstract. A good number of Grid middlewares have been developed recently, yet their effective use for real world applications is seldom validated. Meanwhile, high-performance computing codes keep on requiring more resources than even large clusters can deliver. In this paper, the G-BRAMS project is presented, which aims at executing climatology simulations on a three tiers grid. Three different middlewares (Globus, OAR/CIGRI and OurGrid) are compared and some preliminary measurements are presented. The outcome will be the definition of a Computational Grid, adapted to run a mesoscale meteorological model (BRAMS) used in weather and climatology forecasting that may demand several hundreds of independent executions, one for each different initial condition. The proposed grid is employed to perform these executions for a given middleware. Each execution is scheduled as an MPI job in a particular cluster that is part of the grid.

1 Introduction

The Grid concept was established by Foster [3] as a hardware and software environment with reliable, constant, and cheap access to computer facilities. Several middlewares for grid computing are currently being developed, but few have been used and tested on "real-world" applications. This paper presents an ongoing study of three solutions for Grid Computing, and of their effective usability to run meso-scale meteorological simulations. The three platforms are Globus [4], OurGrid [2] and OAR/CIGRI [1]. Each one of the three groups that participate to the effort is in charge of one specific middleware and offered a cluster in order to test the middlewares in a three tiers grid.[5].

[5] Work sponsored by the FINEP CT-INFO Edital "Grade-01/2004" on the Grid Computing (G-BRAMS project).

The target application is the BRAMS (**B**razilian developments on the **R**egional **A**tmospheric **M**odeling **S**ystem) mesoscale meteorological model used in weather and climatology forecasting. The latter may demand several hundreds of independent executions, one for each different initial condition. This model was developed by the Atmospheric Science Department at the Colorado State University [6, 7] and it has been employed by several universities and research centers in Brazil.

This paper is structured as follows: Sec. 2 presents the scientific context of the project: a climatology study. Next, three sections are devoted to the tested middlewares. Sec 4 gives a qualitative comparison of the three solutions, as well as the first performance measurements obtained on a partial grid running BRAMS. The final section presents our conclusions and future works.

2 Scientific Background: Climatology Simulation

The (B)RAMS [6, 7] model is a multipurpose, numerical prediction model designed to simulate atmospheric circulations spanning in scale from hemispheric scales down to large eddy simulations of the planetary boundary layer. BRAMS is Brazil's most used software for weather and climate prediction, developed and supported by the CPTEC/INPE, and currently installed in some 15 institutions.

CPTEC/INPE is providing global model data for a 10 year climatology. This climatology can thus be obtained using an average over the ensemble, which is a set of long range integration with variations on initial and/or boundary conditions. The International Research Institute for Climate Prediction (IRI) standard strategy has been chosen for the time integration, that is performed on the whole period for each ensemble element. This strategy requires less computational power. The integration domain is divided in three sub-domains, as presented in Fig. 1.

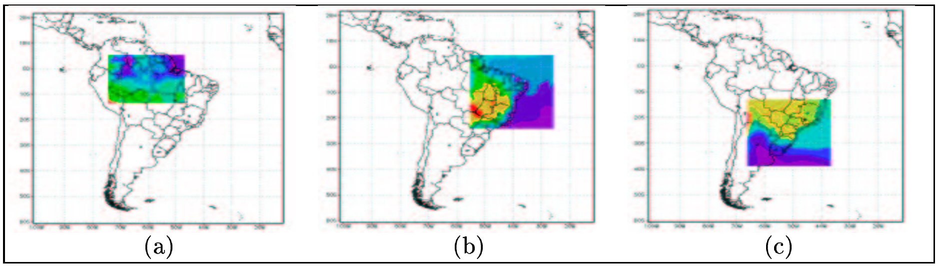


Fig. 1. Space sub-domains: (a) Amazon; (b) North-East; (c) South/Southeast.

The integration will simulate a 10 year climatology, using a checkpoint/restart mechanism that allows to interrupt and resume the integration at every year. For each year, three simulations are performed on each of the three sub-domains, thus providing 9 independent MPI jobs. Each job corresponds to an independent parallel execution of BRAMS.

3 Deploying a Grid for Climatology: Three Middlewares

3.1 Globus

Globus [3, 4] is a set of software tools that implement grid basic services: security, resource management and data transfer. This application employs the following Globus services: user identification (GSI - Grid Security Infrastructure), control and submission of jobs (GRAM - Globus Resource Allocation Manager), and a safe mechanism for data transfer (GridFtp).

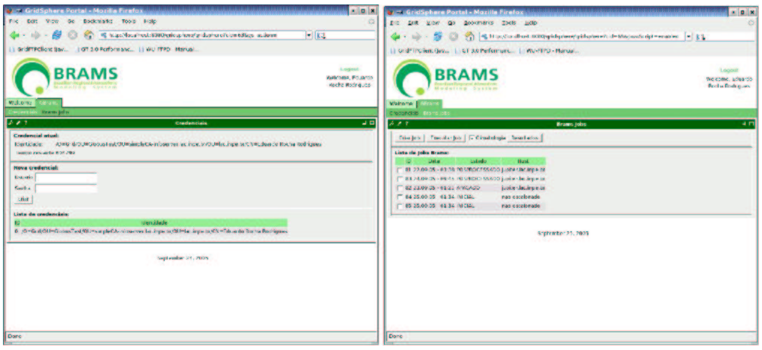


Fig. 2. Globus portal.

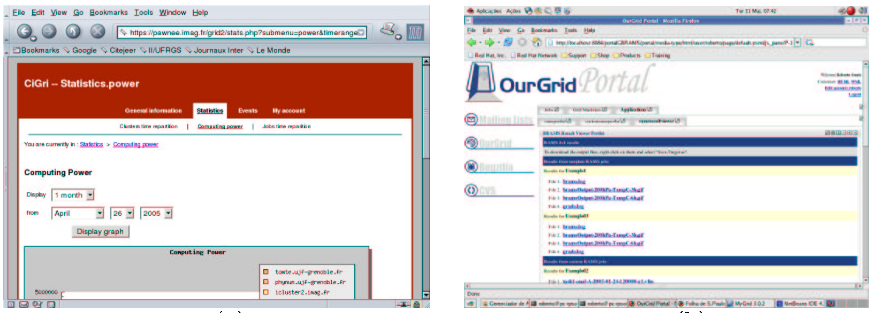
This project uses the Globus toolkit version 3.2. This version does not provide a job scheduler nor load balancing on the grid. A simple scheduler was proposed and implemented based on a queue of jobs and the availability of free grid nodes. A web portal was developed for this Globus application using the GridSphere framework, based on portlets.

3.2 OAR/CIGRI

The CIGRI toolkit [1] is being developed by French INRIA, within the Grid-5000 initiative. It aims at providing a light grid infrastructure, *i.e.* a middleware that does not fulfill all the issues associated with grids, such as security. For the sake of simplicity, CIGRI was developed based on standard off-the-shelf software.

The CIGRI middleware is based on an open-source alternative to PBS (Portable Batch Scheduler) for cluster management, called OAR. OAR makes extensive use of the opensource mySQL database, in order to manage the queue of submitted jobs, as well as the state of the grid nodes. CIGRI is a Perl extension of OAR, that allows to interconnect clusters in a Grid, and to transparently run the tasks in whichever node of the Grid. CIGRI employs tools to rerun failed jobs and to monitor their execution.

CIGRI provides a few other interesting features: a mechanism that regularly collects the job output data on each node, and to gather them on the grid node that shceduled these jobs. Each job can be assigned a priority "weight". CIGRI also includes a php interface that enables the visualization of the grid nodes and the queue of jobs. Fig. 3a shows the current portal of this middleware, based on the version 1.6.



(a) (b)

Fig. 3. (a)CIGRI portal (b) OurGrid portal

3.3 OurGrid

OurGrid [2] is designed for computer grids for applications of the type "Bag of Tasks", *i.e.* fully independent jobs. This middleware links remote machines of different domains and includes three components:

- MyGrid: a user interface for job submission at a home machine;
- OurGrid Peer: it is executed in the peer machine of each administrative domain associated to the grid and provides local machines for job execution;
- User Agent: piece of software installed on each local machine that executes jobs started by the local peer machine.

OurGrid requires Java version 1.4.2 (or superior). The G-BRAMS project involves three administrative domains. Fig 3b shows the OurGrid portal.

4 Comparative Study and Performance Results

4.1 Qualitative Comparison

Table 1 shows some aspects for the three grid solutions. These aspects are related to the installation, management and job scheduling of each middleware.

Due to its "light grid" approach, CIGRI and OurGrid are simpler than Globus, but provide rudimentary scheduling mechanisms. Globus requires the user to implement a job manager or to use a commercial one, but is more robust.

4.2 Performance Results

In order to evaluate the feasibility of each of the three grids, the middlewares have been installed on a cluster composed of 16 Pentium Xeon 2.6 GHz processors, each one with 1 GB of RAM, at CPTEC/INPE and on two other clusters at LAC/INPE and II/UFRGS.

Table 2 shows BRAMS execution times for the CPTEC/INPE cluster, using the data mentioned in Sec. 2 for 1-day and for 10-year simulated time. The latter performs 3 executions per year.

The LAC/INPE and II/UFRGS clusters will be soon replaced by Cray XD1 machines, each one with 12 dual-core processors, reaching 49.6 GFlops/s for the Linpack benchmark. The CPTEC/INPE cluster presents 33.0 GFlops/s estimated for the same benchmark. Based on Table 2, and if the heaviest computation is done by the CPTEC/INPE cluster, the expected average running time of the total simulation, on the final grid, should last around one month.

5 Final Comments

The proposed grid include the use of one of the three middlewares for 3 grid nodes (clusters at LAC/INPE, CPTEC/INPE and II/UFRGS) in order to choose the one that presents the best tradeoff between performance and resource management overheads. The chosen middleware will be then employed in the mesoscale climatology as part of the oprational phase of the G-BRAMS project.

Currently, the three middlewares have been installed and tested on the three grid nodes, concluding the proof of concept of the grid applicability. The presented performance figures of the BRAMS execution show that the climatology is feasible within reasonable execution times in the Xeon and Cray clusters.

It is important to emphasize that each BRAMS job involves a high amount of input and output data, in the order of several of hundreds of Gigabytes.

	Globus	CIGRI	OurGrid
Open source?	YES	YES	YES
Additional S/W	Java	mySQL, Perl	Java
Easy install	NO	YES	YES
Scheduling	user implemented	priority queues	replication
Security	GSI	ssh	ssh

Table 1. Comparison of the 3 middlewares.

Sub-domain	1-day	10-year (3 runs/year)
South-East	310 sec	40 days
North-East	260 sec	33 days
Amazon	355 sec	45 days

Table 2. Execution times of the BRAMS model on the CPTEC/INPE cluster.

Climatology studies required the gathering of such distributed data in order to perform a series of averages of meteorological data obtained in these simulations.

Moreover, the several hundreds of independent BRAMS jobs that are required in a 10-year climatology can be executed using the described machines in a computational grid configuration in a feasible time.

The job scheduler proposed and implemented for the Globus platform provides an inherent load balancing since it tries to schedule 1-year jobs among the grid nodes in a kind of round-robin scheme. The scheduling schemes used on OurGrid and OAR/CIGRI are currently undergoing performance tests.

It is expected that the G-BRAMS project will validate the use of computational grids to allow higher number of simulated years and geographical extension in climatology studies. In addition, the increased computational power of grids may support simulations that couple climatology with other environmental models such as hydrological ones.

References

1. Capit N. *et al.* : A batch scheduler with high level components, Cluster computing and Grid 2005 (CCGrid05), 2005.
2. Cirne W. *et al.* : Building a User-Level Grid for BoT Applications. Book Chapter of High Performance Computing: Paradigm and Infrastructure. Laurence T. Yang, Minyi Guo, editors. John Wiley & Sons Inc., 2005.
3. Foster I. and Kesselman C.: The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1999.
4. Foster I.: Globus Toolkit Version 4: Software for Service-Oriented Systems. IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp 2-13, 2005.
5. Noga J. and Valiron P.: Explicitely Correlated Coupled Cluster r12 Calculations, in "Computational Chemistry: Reviews of Current Trends", 2002.
6. Pielke R.A. *et al.*: A Comprehensive Meteorological Modeling System: RAMS, Meteorological and Atmospheric Physics, 49, 69-91 (1992).
7. Walko R.L., Tremback C.J. and Hertenstein R.F.A : RAMS - The Regional Atmospheric Modeling System, Version 3b - User's Guide, ASTER Division, Fort Collins, CO 80522 (1995).