The SegHidro Experience: Using the Grid to Empower a Hydro-Meteorological Scientific Network

Eliane Araújo   Walfredo Cirne
Gustavo Wagner   Nigini Oliveira
Departamento de Sistemas e Computação, Universidade Federal de Campina Grande
{eliane, walfredo, gustavo, nigini}@dsc.ufcg.edu.br

Enio P. Souza
Departamento de Ciências Atmosféricas, Universidade Federal de Campina Grande
esouza@dca.ufcg.edu.br

Carlos O. Galvão
Departamento de Engenharia Civil, Universidade Federal de Campina Grande
galvao@rechid.ufcg.edu.br

Eduardo Sávio Martins
Fundação Cearense de Meteorologia e Recursos Hídricos
esm9@secrel.com.br

Abstract

This paper describes our experience with SegHidro, a project that empowers hydro-meteorological researchers by (i) enabling collaborative work via the coupling of computer models, and (ii) providing access to massive grid-based computer resources. SegHidro researchers are geographically distributed and have different and sometimes complementary backgrounds. They have in common their interest over the Brazilian Northeast, a semi-arid region, where irregular rainfall distribution causes many problems to the population. SegHidro was created out of the need to help decision makers to better manage water resources. It relies on grid computing as the "glue" that sticks together the community resources: data, computing power and human expertise. Our experience has shown that simplicity is key for the adoption of a solution and its success. Thus SegHidro is built over a simple infrastructure in both computation and data.

1. Introduction

Cheap computing and ubiquitous networking have changed the way we do scientific research, in what has been called e-Science. Computers now play an ever-increasing role in the process of scientific discovery. Data analysis without computers sounds antediluvian. Simulation has joined theory and experimentation as the third scientific methodology. Moreover, grid computing promises to boast e-Science even further by (i) providing access to massive amounts of computational power, (ii) providing remote and easy access to data, and (iii) facilitating scientific collaboration by making it possible for researchers spread throughout the globe to work together in the same effort and/or easily couple together complementary models.

However, this is not to say that e-Science does not carry its own challenges. In fact, realizing the full potential of e-Science face both technical and non-technical obstacles. On the technical side, grid computing technology is new, complex and evolves at a very fast pace. As a result, grid software is typically less stable than one would wish for, and deploying it requires highly specialized skills. Non-technical challenges arise because e-Science changes the way scientists work, thus requiring a new set of skills, rearranging the "power balance" among researchers and institutions, and questioning old values and practices (specially when the grid makes much closer multidisciplinary collaboration possible).

This paper describes the experience and lessons learned trying to overcome such e-Science challenges in the context of the SegHidro project. SegHidro aims to create a cyber-infrastucture to allow scientists and decision makers to better cope with the water problem of the Brazilian northeastern region, enabling them to better manage the water resources. The Brazilian Northeast is a semi-arid region with a very irregular rainfall distribution. As such, people suffer with droughts, as well as with eventual flooding. In fact, during the 20th Century, drought events have caused losses of lives, mainly on the poor population that lack drinking water, food and work [20]. The rainfall irregularity makes it especially hard to manage the region's water resources. For example, after a good rainy season, should people use the water in the reservoirs for massive irrigation, or should they save water to assure city supply in the next dry seasons?

SegHidro intends to foster the understanding of the water problem by allowing scientists to collaborate and refine their hydro-meteorological forecasting models. And also to improve water management by making good models easily available to the decision makers. In order to achieve this goal, we rely on grid computing to allow geographically distributed teams to share computing power and the data produced by them. In fact, SegHidro integrates the models produced by researchers with different backgrounds, enabling them to communicate via their software, not only through natural language. Coupled with the greater computational power offered by grids, this arrangement is taking the scientific cooperation to a whole new level. In our case, the whole is truly greater than the sum of the parts.

As we shall see in details, the main lesson learned is that *simplicity* is the key for having the system being truly used in practice. Application scientists and decision makers should not have to invest much time learning new tools or radically changing the working routine. In particular, scientists should be able to run their applications on the grid by themselves, with minimal training and support from grid developers.

This paper is structured as follows. Section 2 details SegHidro project. Section 3 is about the cyber infrastructure that this project relies on; it presents the grid architecture and SegHidro core architecture. Section 4 points some related works. Section 5 concludes the paper and presents some future works.

2. SegHidro project

SegHidro stands for "Segurança Hídrica" (Hydrological Security) and aims at improving the water management of the Brazilian Northeast, a semi-arid region. We hope to provide better tools for the decision makers that oversee our water supplies, enabling them to reduce the risk of their decisions while minimizing the effects of water scarcity in our region. We also hope to raise the transparency of the decisions on the water supply by making the technical rationale behind the decisions public and openly discussed. Note that this is possible because the data collected and analyzed by SegHidro members are essentially public. Indeed, the more public that data truly becomes, the better for the discussion and understanding on water management. Besides the scientists, SegHidro must support also a number of technical and political people to make and implement water-related decisions at the various levels of decision (from the nation-wide allocation of agricultural credit lines, for example, to the management of the small reservoir that provides water to a rural village). As such, SegHidro is open to the general public, as one can witness in the project portal at http://portalseghidro.dca.ufcg.edu.br/.

In alignment with the project goals, openness and decentralization have been fundamental SegHidro values since the beginning of the project. SegHidro is not centralized in any way, shape or form. It has no owner. It is a federation that scientists and decision makers join because it is in their own interest to do so. Note that this approach leads to an organic growth of the grid infrastructure, both in the technical capabilities it supports, as well as in the institutions and people that participate in the grid.

In fact, SegHidro is being built over an already existing scientific network that congregates atmospheric and environmental scientists that, for years, have been conducting joint projects, regularly meeting, exchanging experiences, and discussing water-related issues in the Brazilian Northeast. This community then incorporated computer scientists to build and operate SegHidro, a grid that supports and fosters its activities.

SegHidro uses OurGrid as its grid middleware [24]. While OurGrid provides an open and totally decentralized compute infrastructure, SegHidro itself must establish minimal rules to guarantee that the applications from different researchers can work together.

A typical hydro-meteorological application is a cascade of models that simulates the behavior of the atmosphere, hydrographic basins, aquifers and reservoirs. These models are computer programs that reproduce physical, chemical and biological processes of the earth system, such as meteorological and hydrological ones.

In particular, atmospheric models, that perform climatic and weather forecasting, require a high volume of input data, a tremendous computational power and produce a huge amount of data. These results are the primary input of the other models on the cascade.

The global circulation models (GCM), used to generate climate predictions, have a typical horizontal spatial scale (larger than 100 km) that overlooks some local forcing such as topography, sea-land contrast and land use [25]. On the other hand, regional modeling provides higher resolution forecasts for a limited part of the globe. The technique called dynamical downscaling allows one to use a meso-scale regional model to enhance the information already generated by a GCM. 'Nesting' a regional model into a GCM means to run it with initial and boundary conditions from the GCM. Since atmospheric models possess several possibilities of dealing with physical processes such as radiation, clouds, turbulence, etc, each model setting can be considered as a particular model and the corresponding simulation is considered as a member of the forecast. Numerous forecasts members can also be produced from a particular global-regional coupled model by varying certain initial conditions [10]. This so-called ensemble methodology assures a more accurate characterization of the uncertainty associated to the forecasts. When several different coupled models are used, a super-ensemble of forecasts is produced. Since the models have different forecast skills for different atmospheric settings, weighted combined forecasts are very useful for actual decision making.

Rainfall and temperatures forecasts provided by the atmospheric models are input for hydrological, hydrogeological and agricultural models, among others. The resulting simulations are used by management models, such as for operation of water supply reservoirs and distribution pipeline networks, flood control, aquifer exploitation, irrigation scheduling and early warning. All these models are site-specific and, as for the atmospheric models, can have multiple parameterizations, which, in the end, generate a simulation tree with several scenarios for a particular site. Regional water resources and agricultural management centers are interested in forecasts for hundreds of sites in their geographical area of responsibility.

Figure 1 exemplifies an output of processing three of such models for a river basin in northeastern Brazil [4]. It presents seasonal (January–June) simulations of precipitation mean daily discharge and reservoir storage volumes at the end of June of each year, using a cascade of atmospheric, hydrologic and reservoir simulation mode. For simplicity, only three members of the ensemble are shown ('forecast 1', 'forecast 2' and 'forecast 3'), their average and the actual observed values. This kind of experiment, using a long time series of historical observations (1971-1988 in this case) is usually performed to evaluate models' skill in representing the processes under simulation. This particular experiment was carried out by a team of researchers on meteorology, hydrology and water management from six research centers and universities.

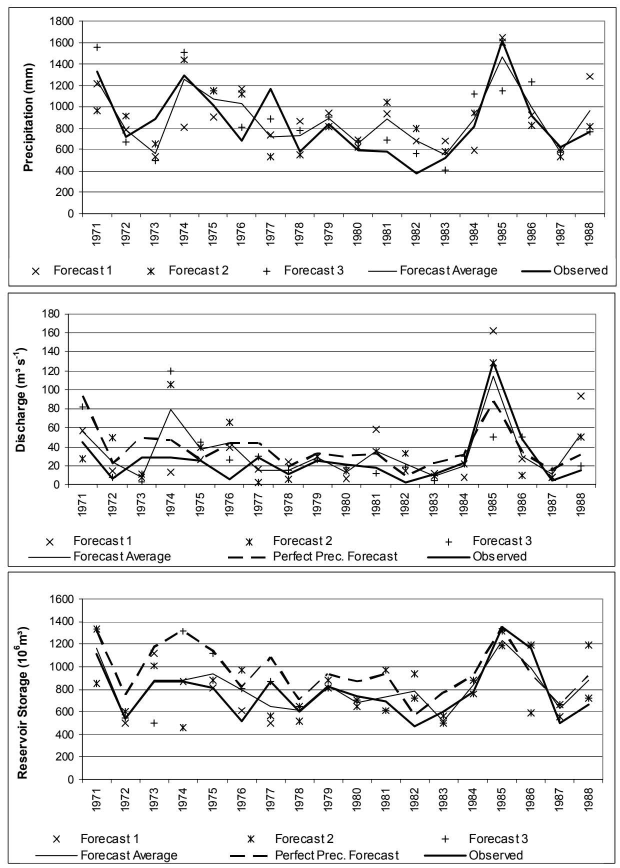

Figure 1 - Piancó river basin in Brazil

While GCMs run on supercomputers at national and international meteorological centers, regional models can run in single workstations or clusters, provided the necessary initial conditions from the GCM are made available. Hydrological, hydrogeological and agricultural models are less computer-demanding, but management decision models can be high-demanding. Present limitations in computing power of several Stated-based forecasting centers in Brazil often lead to running small ensembles or even a single deterministic run of the cascade of models.

For this reason, SegHidro allows running the simulation tree generated by the ensemble methodology. It is easy to conclude that the need of computing power and the parallel execution is great, but enabled by the grid environment.

The rules established by SegHidro enable a very effective collaboration among people with different background in the sense that models can be easily coupled. In fact, we decided for a very simple file-based text format to enable the interoperation of the models. A file-based format was chosen because it is well understood and easily implemented by non-computer scientists, most of whom use Fortran as programming language. We have defined formats to enable the coupling from atmospheric models (both for climate and weather forecasting) to agriculture or hydrological models, and those to reservoir models (both surface and underground).

The scientific network underlying SegHidro includes experts on each model of the cascade. Their interaction is important in order to allow data interoperability among models and, consequently, reuse of data generated by different teams. As a result, SegHidro enables its participants to do new science, which was difficult, less efficient, or just impossible without it. It also greatly eases experimentation and comparisons, since it makes straightforward to replaced a model by another one with similar functionality. More then a dozen of institutions are part of SegHidro effort. They are regional and national meteorological production centers, as well as universities and research centers. They provide critical information to governments agencies, civil defense, the press and so on.

3. Cyber infrastructure

The cyber infrastructure constructed by SegHidro relies on OurGrid, that allows sharing resources and data, and on its core architecture that establishes data standards and the way the models cascade would behave.

3.1 Grid Architecture

OurGrid is an open, free-to-join, cooperative grid in which institutions donate their idle computational resources in exchange for accessing someone else's idle resources when needed [13]. By free-to-join, we mean that anyone can download the OurGrid software from http://www.ourgrid.org and join the grid. There is no paperwork or human negotiation regarding what each institution, which we will call site, contributes to and receives from the grid. OurGrid uses the Network of Favors, a peer-to-peer technology that makes it in each site's best interest to collaborate with the system by donating its idle resources. OurGrid leverages from the fact that people do not use their computers all the time. OurGrid strives to be non-intrusive, in the sense that a local user always has priority for local resources. In fact, the submission of a local job kills any foreign jobs that are running locally. This rule assures that OurGrid cannot worsen local performance, a property that has long been identified as key for the success of resource-harvesting systems [6]. OurGrid can use both interactive desktop computers and dedicated clusters (which may be controlled by a resource manager, such as Maui [5], PBS [1], and CRONO [12].

In OurGrid, each site corresponds to a peer in the system. A given peer in OurGrid will commonly run tasks from other unknown peer that is also part of the community. This creates a very obvious security threat, especially in these days of so many software vulnerabilities. Therefore, we must provide a way to protect local resources from foreign unknown code. That is the job of SWAN (Sandboxing Without A Name), a solution based on the Xen virtual machine [18], that isolates the foreign code into a sandbox, where it cannot access local data nor use the network. Naturally, we must also protect the application from malicious peers. We do that with low overhead by using the credibility-based sabotage detection proposed by Sarmenta [18].

For now, at least, OurGrid assumes applications to be Bag-of-Tasks (BoT), those parallel applications whose tasks are independent. However, a single OurGrid task may itself be a parallel tightly-coupled application (written in MPI, for example). Although OurGrid does not run parallel tasks across the grid, it may very well run them on a remote site. Since decision makers typically run parameter sweeps to evaluate the risk of their decisions, support for BoT-only applications is adequate for SegHidro.

Users and applications may interact with OurGrid via MyGrid, a personal broker that performs application-level scheduling and provides a set of abstractions that hide the grid heterogeneity from the user.

In summary, OurGrid has three main components: the OurGrid peer, the MyGrid broker, and the SWAN security service. Figure 1 shows them all, depicting the OurGrid architecture. OurGrid is in production since December 2004 and now encompasses more than 20 sites that together have around 400 machines. Six of these sites are SegHidro members. OurGrid current status is available at http://status.ourgrid.org/.

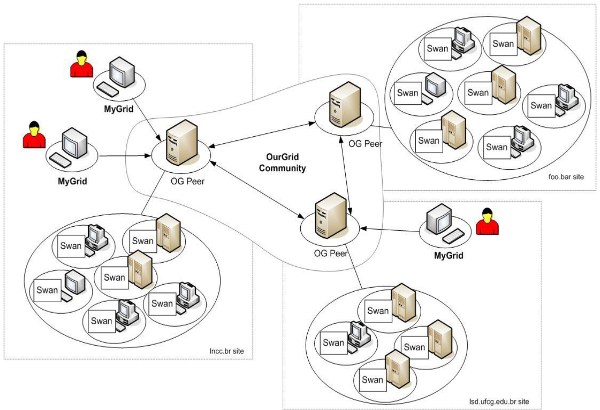

Figure 1 - OurGrid Architecture

3.2 SegHidro Core Architecture

SegHidro Core Architecture, also known as SHICA, is the computing infrastructure to support the execution of SegHidro applications. SHICA interacts with MyGrid in order to submit jobs to be executed on the grid and gather their results. At SegHidro, a web-portal was developed to be an entry point to SHICA and the grid. However, it is not an exclusive alternative. Standalone applications can directly access SHICA as well. This is important to users who are refining the models and need to run parameter sweep and/or statistical evaluation of their changes. For those users, the ability to write scripts is paramount.

SegHidro applications, those hydro-meteorological applications, can be viewed as a workflow. We consider a flow to be a complete sequence of a SegHidro application, as can be seen on Figure 2 Each part of the flow is a model, and the output of a preceding model serves as input to its succeeding model. The models follow a precedence rule on which atmospheric models are the base, followed by agriculture or hydrological model, and, optionally, reservoirs management models. Note that each model on the flow can represent different instances, i.e. atmospheric models may be: RAMS [21], RSM [9], ECHAM [7], etc.

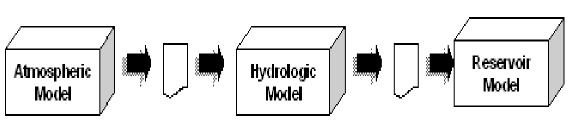

Figure 2 - SegHidro Flow

Data Format. We have analyzed the requirements of different models and have defined a very simple file-based format with which those different models instances can interoperate. This approach eases the process of entering a new model to the system or substituting one. The format convention was discussed with SegHidro community in order to assure the adoption of the system by the institutions and researchers.

Each model that accomplishes to SegHidro must receive or return data in the specified file-based format. These formats are known by SegHidro community as PMH (atmospheric model output) and PHR (hydrological model output). A sample of PMH file is shown at Table 1, where X = latitude; Y = longitude; E = elevation; D = date; H = time; P = precipitation; T = temperature. For sake of space we are omitting PHR file format.

Table 1 - PMH File Format

!X	Y	E	D	H	P	T
-40.3500	-6.3500	57	2005-01-01	00:00:00	0.0	-
-40.3500	-6.3500	57	2005-01-02	00:00:00	5.5	-
-40.3500	-6.3500	57	2005-01-03	00:00:00	4.0	-
-40.3500	-6.3500	57	2005-01-04	00:00:00	4.0	-
-40.3500	-6.3500	57	2005-01-05	00:00:00	2.7	-

Before defining this data model, other file formats were evaluated such as NetCDF [22] and HDF [14]. They are standards to describe scientific data, especially from environmental researches. Both are complete, extensible and versatile but, much more complicated than SegHidro applications needs. SegHidro project aims to empower its community to produce better science and anything that would be a barrier to use the system must be discarded. The PMH file, created in a common agreement, was defined to have the minimal information from the atmospheric model needed to the subsequent models. The community demanded a solution as simple as possible, since they need to adapt their models to it.

This approach benefits simplicity and collaboration. For example, anyone who needs forecasting input data to execute their hydrologic model first need adjust it to receive data in the SegHidro format, then he or she can use data generated from whichever atmospheric model available in the system. Furthermore, adjusting the hydrological model output to adhere PHR file format, enables the execution of whichever reservoir management models available in the portal (provided the user inform the requested input data).

Data Access. In order to access data produced by models at SegHidro system, many approaches would be adopted in the trend of web services such as WS-Reliable Messaging [16] or Open grid services data access integration [17]. OGSA-DAI is a middleware that allows data integration from remote sources in a grid environment. Another interesting alternative seems to be OPeNDAP, a framework that makes it easy scientific data exchange [19]. This project was born from the

Distributed Oceanographic Data System (DODS), a widely accepted project in the oceanography and meteorology areas. OPeNDAP is flexible in terms of its data model and may solve some problems we foresee regarding to atmospheric model output data slicing. We are evaluating integration with SHICA and OPeNDAP data sources. By now, the current implementation access data through Java-RMI connections.

Data Discover. It is highly desirable that data obtained from the atmospheric models could be reused. For instance: one has executed the climatic forecast for whole Brazilian Northeast, as said before, a process that demands high computing power. The computation may take nine hours in a dedicated 20-processors cluster and produces, approximately, four Gigabytes of data for each member. Someone else needs to execute a hydrological model over a basin of Pernambuco. Since Pernambuco is a state of Brazilian Northeast, we would like to use the previously generated forecast without having to re-execute it.

The problem here is how to discover that someone has already executed the desired simulation; where the output result is and how to access it. In data grid research domain [2], a classical solution is to use a catalog, which can be viewed as a map that relates the physical resource name to a logic name. This logic name must distinguish one replica (output data) to another. Currently SHICA implementation accesses a single centralized catalog.

Many issues arose when dealing with data as a resource on grid, such as: how to deal with wrong information from the catalog; how to schedule considering data location (move computing to data); how to efficiently maintain and update a catalog and, more related to OurGrid issues, how to account data favors on the network of favors [24].

OurGrid storage affinity scheduling heuristic [8] was conceived to deal with huge amount of data. The rationale behind storage affinity is to, a priori, detect which site is more amendable to execute a given task and then schedule the task to that site. The affinity of task and site is calculated in terms of how many bytes need to be transferred in order to the task to start. The less bytes to transfer, the bigger the affinity. On SegHidro perspective, storage affinity search on the catalog which site stores the desired data so that the computation is moved to that site.

Regarding the other issues, we do not have definitive answers yet. As data grid is a relatively new research domain, our team is studying, discussing and developing alternatives. However, since grid computing world is moving towards service oriented architecture [15], in order to ease the development and composition of the systems, we seem that encapsulating some of these needs in services would be a good approach to tackle with the problems.

4. Related Work

There are other initiatives that, similarly to SegHidro, intend to provide an e-science infrastructure to environmental researches such as LEAD [10] and ESG [3]. Obviously they focus on their particular needs and requirements. However, it is worthwhile to look at them, since they are experienced projects which certainly will teach us some lessons. Furthermore, as the grid middleware is fundamental to our system, it is important to analyze other alternatives such as GridBus [23].

4.1 LEAD – Linked Environment for Atmospheric Discover

LEAD is an ambitious American project that aims to create a scalable cyber-infrastructure, based on grid computing, to help researchers and operational practitioners to do meteorology research and education. They focus on numerical weather prediction running the Weather Research Forecast (WRP) model. They claim to provide the community "with an easier way to perform regionalized mesoscale model runs" [10]. LEAD encompasses a set of tools to perform data assimilation, preparation, analysis, visualization and so on.

The draw of LEAD is its on-demand and adaptive features. Similarly to SegHidro, the executions are modeled as a workflow. However, on LEAD, processes initiates automatically and rapidly may change its configuration in response to the weather. This project is an on-going work and its development is based on service oriented architecture (SOA). The set of services allows the workflow composition on-the-fly. A web portal is the system front-end such as SegHidro.

4.2 ESG – Earth System Grid

The ESG project targeted their efforts toward the problems faced by the community climate modeling. In order to execute their models, it is necessary high computing power and an archival area able to store hundreds of Giga or some Terabytes of output data. Furthermore, in ESG context, for the user to consume this data it is necessary to move it to the user's desktop. ESG prototype development was based on some data grid technologies [3], such as: replica management, data transfer, request management and so on.

Data transfer is a central issue on ESG system and it is performed by GridFTP. SegHidro system also needs to deal with this problem; however the approach is to diminish data transfers by moving computation to data.

4.3 GridBus

The Gridbus [23] extends the computational grid resource broker to a data-oriented broker. As e-science applications demand transfer of large amount of data to computational resources, it is mandatory to optimize application scheduling to achieve efficient performance on data-intensive applications. Gridbus scheduler take into account the time that a machine will compute a job analyzing past executions on it, the data distance and the bandwidth between machine and data host.

SegHidro adopts a similar approach using storage affinity [8] heuristic. The objective here is to schedule the jobs to be executed in a set of machines with a better affinity, which means the amount of input data necessary to be transferred in order to execute the job. Furthermore, data reuse supported by SHICA contributes to decrease the need of data transfers.

5. Concluding remarks

In this paper we presented our experience on building SegHidro and analyzed how it empowers hydro-meteorological who have common interests, namely the water issue in the Brazilian Northeast. SegHidro leverages collaboration in many aspects: computational power, data and human expertise.

A SegHidro application is a workflow of scientific simulation models, as such, one model's output serve as input to its succeeding, which is called model coupling. The system makes it easier to researchers execute their simulations by adjusting input parameters or replacing models, enabling scientists to collaborate through their models, creating a combined result that truly is the sum of their (complementary) expertise.

We believe two aspects of SegHidro were critical to achieve such results. First, the people involved in SegHidro have a common interest. Therefore, it makes sense for them to share experiences and collaborate in the application level, sharing much more than just computational power. Second, we have kept the system as simple as possible, making it feasible for busy application scientists to effectively use it. A key part of such simplicity relates to recognizing that no single interface would fit the needs of everybody. We thus have a web portal for those who use the system more in production mode in their decision making, as well as stand-alone script-friendly applications that enable scientists working on the models to efficiently carry on their work.

In short, the main contribution of SegHidro was to create an infrastructure and its rules that fosters collaboration among researchers and decision makers: meteorologists produces their forecasts according to their studies, hydrologists consume them according to their needs, decision makers can simulate many scenarios to devise a better choice using their models. Further, the system makes it ease to evaluate how one's "piece of the whole" contribute to the final result, providing thus invaluable feedback for the researchers on to where refine their models.

6. References

[1] A. Bayucan. Portable Batch System Administration. Veridian System, 2000.

[2] A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke. The data grid: Towards architecture for the distributed management and analysis of large scientific datasets. Journal of Network and Computer Applications.

[3] B. Allcock, I. Foster, V. Nefedova, A. Chevrenak, E. Deelman, C. Kesselman, et al. High-Performance Remote Access to Climate Simulation Data: A Challenge Problem for Data Grid Technologies. In proceedings of SC' 01.

[4] C. O. Galvão, P. Nobre, A. C. F. M. Braga, K. F. Oliveira, R. M., S. R. Silva, M. F. Gomes Filho, C. A. G. Santos, F. Lacerda, D. Moncunill, 2005. Climatic Predictability, Hydrology and Water Resources over Nordeste Brazil. In: Regional Hydrological Impacts of Climatic Change — Impact Assessment and Decision Making. Wallingford, International Association of Hydrological Sciences Press.

[5] D. B. Jackson, Q. Snell, and M. J. Clement. Core Algorithms of the Maui Scheduler. In 7th International Workshop, JSSPP 2001, Cambridge, MA, USA, 2001.

[6] D. Thain, T. Tannenbaum, M. Livny. Distributed Computing in Practice: The Condor Experience. Concurrency and Computation: Practice and Experience, volume 17, issue 2-4, pages 323-356, February-April 2005.

[7] E. Roeckner. Simulation of the present-day climate with the ECHAM model: impact of model physics and resolution. Publisher: Max-Planck-Institut für Meteorologie, 1992

[8] E. Santos-Neto, W. Cirne, F. Brasileiro, A. Lima. Exploiting Replication and Data Reuse to Efficiently Schedule Data-intensive Applications on Grids. 10th Workshop on Job Scheduling Strategies for Parallel Processing, June 2004.

[9] H. M. H. Juang, M. Kanamitsu. The NMC nested regional spectral model. Monthly Weather Review, 122, 1994

[10] K. Droegemeier, V. Chandrasekar, C. Richard et al. Linked environments for atmospheric discovery (LEAD): Architecture, Technology Roadmap and Deployment Strategy, 21st Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology, January 2005

[11] L. Sarmenta. Sabotage-Tolerance Mechanisms for Volunteer Computing Systems. Future Generation Computer Systems, Vol. 18, Issue 4, Elsevier, 2002.

[12] M. A. S. Netto and C. A. F. de Rose. CRONO: A configurable and easy to maintain resource manager optimized for small and mid-size GNU/Linux Clusters, October 2003.

[13] N. Andrade, F. Brasileiro, W. Cirne, M. Mowbray. Discouraging Free-riding in a Peer-to-Peer CPU-Sharing Grid. 13th IEEE International Symposium on High-Performance Distributed Computing (HPDC13), June 2004.

[14] National Centre for Supercomputing Applications HDF5 - A New Generation of HDF. Accessed at August, 2005: http://hdf.ncsa.uiuc.edu/HDF5/

[15] OASIS – Service Oriented Architecture Reference Model Technical Committee. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=soa-rm Accessed at August 2005.

[16] OASIS – Web Service Reliable Messaging (WSRM) Technical Committee. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsrm Accessed at August, 2005.

[17] Open Grid Services Architecture Data Access and Integration Specification. http://www.ogsadai.org.uk/ index.php Accessed at August, 2005.

[18] P. Barham et al. Xen and the Art of Virtualization. SOPS'2003.

[19] P. Cornillon, J. Gallagher and T. Sgouros, "OPeNDAP: Accessing Data in a Distributed, Heterogeneous Environment", Data Science Journal, 2, November 2003, pp 164-174

[20] P. Kabat, R.E. Schulze, M.E. Hellmuth, J.A. Veraart (editors), 2002. Coping with impacts of climate variability and climate change in water management: a scoping paper. DWC-Report no. DWCSSO-01(2002), International Secretariat of the Dialogue on Water and Climate, Wageningen.

[21] R. A. Pielke, W. R. Cotton, R. L. Walko, C. J. Tremback, W. A. Lyons, L. D. Grasso, M. E. Nicholls, M. D. Moran, D. A. Wesley, T. J. Lee, and J. H. Copeland, 1992: A comprehensive meteorological modeling system--RAMS. Meteorology and Atmospheric Physics, 49.

[22] R. Rew, G. Davis and S. Emmerson. NetCDF Homepage http://my.unidata.ucar.edu/content/software/netcdf/index.html

[23] S. Venugopal, R. Buyya, and L. Winton, 2004. A grid service broker for scheduling distributed data-oriented applications on global grids. In Proceedings of the 2nd Workshop on Middleware For Grid Computing (Toronto, Ontario, Canada, October 18 - 22, 2004)

[24] W. Cirne, F. Brasileiro, N. Andrade, R. Santos, A. Andrade, R. Novaes, M. Mowbray. Labs of the World, Unite!!! Universidade Federal de Campina Grande, Departamento de Sistemas e Computação, UFCG/DSC Technical Report 01/2006 Submitted for publication.

[25] Y. Wang, L. R. Leung, J.L. McGregor, D.-L. Lee, W.-C. Wang, Y. Ding and F.Kimura, 2004: Regional climate modeling: progress, challenges and prospects. Journal of the Meteorological Society of Japan