CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. 2007; **19**:1609–1620
Published online 7 June 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cpe.1118

Fostering collaboration to better manage water resources

William Voorsluys[1], Eliane Araújo[1], Walfredo Cirne[1,2,*,†], Carlos O. Galvão[3], Enio P. Souza[4] and Enilson P. Cavalcanti[4]

[1]*Laboratório de Sistemas Distribuídos, Universidade Federal de Campina Grande, Bodocongó, Campina Grande, PB, Brazil*
[2]*Departamento de Sistemas e Computação, Universidade Federal de Campina Grande, Bodocongó, Campina Grande, PB, Brazil*
[3]*Departamento de Engenharia Civil, Universidade Federal de Campina Grande, Bodocongó, Campina Grande, PB, Brazil*
[4]*Departamento de Ciências Atmosféricas, Universidade Federal de Campina Grande, Bodocongó, Campina Grande, PB, Brazil*

SUMMARY

Good water management is literally vital for the arid and semi-arid regions of the planet. Yet good water management requires multidisciplinary expertise, since one must consider climatic, hydrological, economical and social aspects to make balanced decisions on water usage. We here present SegHidro, a Grid portal designed to foster scientific, technical and operational collaboration to improve water resources management. The portal targets researchers and decision makers, enabling them to execute and couple their computational models in a workflow. The portal provides a framework which allows seamless integration of the models, meaning that each phase of the flow may be executed by a different expert and that the resulting data are shared among other portal users. Due to the nature of these applications and the need to execute many prospective scenarios, their execution requires high computing power. However, we go beyond providing high-performance computational Grid capabilities. We also enable people to complement each other's expertise in understanding the trade-offs in the water allocation decisions. The SegHidro portal is about sharing: human expertise, data and computing power. Copyright © 2007 John Wiley & Sons, Ltd.

Received 6 December 2005; Accepted 6 July 2006

KEY WORDS: distributed systems; Grid computing; Grid portal; e-Science; water management

*Correspondence to: Walfredo Cirne, Laboratório de Sistemas Distribuídos, Universidade Federal de Campina Grande, Av. Aprígio Veloso, 882—Bloco CO, Bodocongó, Campina Grande, PB, CEP 58109-970, Brazil.
†E-mail: walfredo@dsc.ufcg.edu.br

Contract/grant sponsor: Brazilian Ministry of Science and Technology (FINEP/CT-INFO); contract/grant number: 01.04.08888.00

1. INTRODUCTION

Grid computing has emerged to provide an infrastructure to execute high-performance computing applications empowering users with resources they can share or pay for. At the moment, however, the vision of a pervasive and ubiquitous Grid is not fully implemented [1]. The Grid infrastructure congregates machines distributed geographically and administratively into a single virtual parallel computer, multiplying the computational power and diminishing application execution time. The orchestration of distributed and heterogeneous resources, along with security requirements, makes the Grid a complex environment to set-up and use. It is difficult to assemble Grids, acquire machines and fulfill the security requirements imposed by network administrators. These problems scare users and institutions, preventing them from using the Grid to perform their work [2]. Grid portals help scientific users to execute their application in a Grid environment. In general, they can be generic, providing access to Grid resources and services, or specific, enabling users to run a previously configured application in a Grid infrastructure, which remains transparent to the user. A Grid portal is therefore a problem-solving environment that allows users to develop, access and execute distributed Grid applications from a conventional Web browser. It is the user point of access to a complex Grid system [2–4].

The SegHidro portal, presented in this paper, is being developed by the SegHidro project and can be reached at http://portalseghidro.dca.ufcg.edu.br. The project aims to allow scientists and decision makers to better cope with the water problem of the Brazilian northeast region. The irregular rainfall distribution of this region has caused, over the centuries, problems to the population, especially to the poor people. Long periods of drought events result in damages to agriculture and people water supply. On the other hand, the coastal area, where the largest cities of the region are located, is subject to tropical storms and floods. The portal provides a means to simulate a wide variety of scenarios, based on weather and climate forecasts, in order to better decide about water reservoir management, agricultural planning and flood control. These simulations are performed using computational models developed by meteorologists, hydrologists and engineers. Their results help to predict behaviors that are fundamental inputs on water management decisions. As we need to simulate many scenarios and weather and climate forecasts are very resource demanding, the computational resources provided by the Grid are paramount.

The portal embraces a framework that allows model coupling. A typical SegHidro application is a workflow in which the input of a model is the output of its precedent. Portal users can upload their own models or use existing models previously inserted by other users. The output data generated by the executions may be shared by the users. Scientists and model developers use the portal as a vehicle to make available and test their models. The models (or forecasts generated by them) stored and available at the portal are commented on by their developers, so that they can be safely used by other users without necessarily having the knowledge of the modeling procedures. For example, hydrologists can use climate forecasting information produced by the meteorologists for generating drought occurrence probabilities, with confidence and very easily. These data are dispersed on the Grid and can be retrieved through a replica location service. This feature is implemented by the portal and is transparent to the users.

The gist of SegHidro portal is to provide a unique platform to allow people with different backgrounds to collaborate and work together. The platform objectives go beyond providing high-performance computing tools to the users. The collaboration aspects of the platform are especially relevant. We present in this article how this platform and the Grid are changing the way governmental agencies, production centers and final users work on their activities related to water management.

This paper is structured as follows. Section 2 details SegHidro portal architecture and Grid middleware. Section 3 discusses the collaboration aspects of the portal and computational empowerment. Section 4 presents two cases where SegHidro Portal is useful. Section 5 discusses the related work. Finally, Section 6 concludes the paper with some final remarks and points future portal enhancements.

2. SEGHIDRO PORTAL

The SegHidro portal is the means of interaction with scientists and decision makers who wish to run simulations in production mode. The portal allows users to upload their own models through a flexible and easy-to-use interface. Once inserted into the system, the model becomes available to every portal user who wants it to take part in a workflow execution. The workflow approach allows model coupling, a technique where the output of one model serves as the input to the succeeding model. In this current version, the flow may be composed by two or three levels.

The first level is reserved for regional meso-scale numerical atmospheric models, which provide short-term weather forecasts as well as long-term climatic predictions. Regional models depend on the assimilation of predictions generated by global circulation models (GCMs) [5]. Our portal makes it possible to use GCM results produced by different centers. Atmospheric models possess several possibilities of dealing with physical processes and each run may be produced from a particular global–regional coupled model. Each model configuration can be considered as a particular model and the corresponding simulation is considered as a member of the forecast. A set of numerous forecast members is called an ensemble. The technique of using a set of members to assure a more accurate characterization of the uncertainty associated to the forecasts is called an ensemble methodology. 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. In the portal, the ensemble obtained in the first flow level is extended to the next levels. Rainfall and temperature forecasts provided by the atmospheric models are input for hydrological, hydro-geological and agricultural models, which reside on the second level of the flow. The third level of the flow is reserved for management models, such as operation of water supply reservoirs. These models are site-specific and, as for the atmospheric models, can have multiple parameterizations, which, in the end, generate a parameter sweep 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 illustrates how the models are coupled and how different instances of the models may produce several distinct execution flows.

Note that the presence of multiple possibilities in each level of the workflow, motivated by the ensemble, leads to a high demand for processing power and raises concerns about the management of the data involved. In particular, the atmospheric models are highly computationally intensive and produce massive amounts of output data. These requirements demanded a significant effort when building the computing infrastructure to back up the SegHidro portal. We have applied several technologies: some are well established and some have been specifically designed to support the project. The software technologies employed may be divided into two categories: (1) those used to build the portal interface and (2) the underlying technology, which forms the support layer that allows the execution of the applications in an efficient and transparent way. To build the user interface,

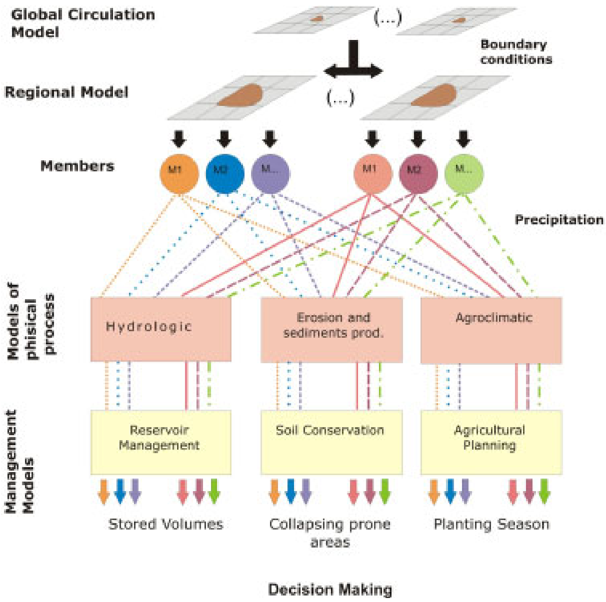

Figure 1. Model coupling at the SegHidro portal.

we have used the Jetspeed framework along with other Web technologies such as JSP, servlets, HTML and JavaScript. Several tools reside on the lower layer, including: the OurGrid middleware; SHiCA, a broker that manages the workflow construction and its execution through on the Grid; and, finally, a data management layer, which provides a means to efficiently cope with many operations regarding distributed data.

2.1. User interface

The Web portal interface has been built over the Jetspeed framework. Jetspeed [6] is an open-source portal framework provided by Apache. It allows the creation of portlets that may be easily integrated and deployed into the portal. The framework has built-in services for user interface customization, caching, persistence and user authentication. This framework allowed the development team to focus on the creation of the application itself, bringing significantly faster results.

Five portlets have been created: (1) meteorology, (2) model editing, (3) flow execution, (4) results and (5) jobs. The 'meteorology' portlet has been conceived to allow users to execute a pre-installed atmospheric model; users may easily adjust some model parameters, select the desirable variables and configure the simulation period. Through the 'model editing' portlet, users may upload and configure new models in the portal; these models are typically hydrological and reservoir management models. In the 'flow execution' portlet users may compose a workflow with the uploaded models and submit it for execution on the Grid. The 'results' portlet is used to retrieve the results of each model in

DOI: 10.1002/cpe

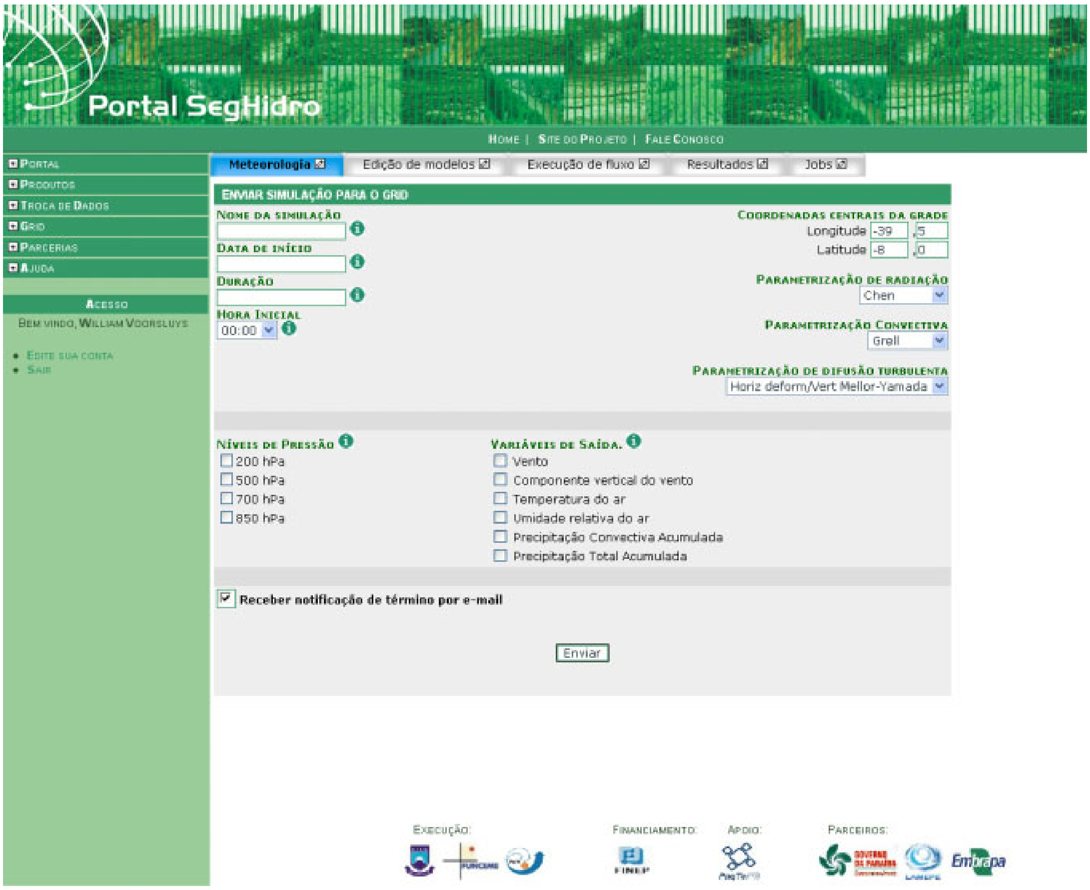

Figure 2. A screenshot of the portal interface.

the workflow. The portlet named ‘jobs’ is a monitoring tool that allows users to verify the overall status of the submitted jobs. Figure 2 shows a screenshot of the portal interface with the focus on the ‘meteorology’ portlet.

The interface used to add new models is generic enough to allow the insertion of any model, provided that it is prepared to receive the outputs of preceding models, if any, in the standard formats. Also, the program must be executed in a non-interactive manner. The system assumes that a model has a name, a type, an executable file, a command line with parameters, a set of input files or scripts, a list of output file names and a textual description. The type determines at which position, in the workflow, the model will fit. All parameters registered may be adjusted when the user is composing the workflow. Only the user that has uploaded the model is allowed to make persistent changes to it.

DOI: 10.1002/cpe

2.2. OurGrid

The high-performance computing architecture constructed by SegHidro relies on OurGrid, which 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 [7,8]. 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 a site, contributes to and receives from the Grid. OurGrid leverages from the fact that people do not use their computers all of the time. It 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.

It 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 to the success of resource-harvesting systems [9]. OurGrid can use both interactive desktop computers and dedicated clusters (which may be controlled by a resource manager, such as CRONO [10]). For now, at least, OurGrid assumes that applications are Bag-of-Tasks (BoT), 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.

In OurGrid, each site corresponds to a peer in the system. A given peer will commonly run tasks from another unknown peer that is also part of the community. This creates a very obvious security threat, especially given the present level of 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 [11], that isolates the foreign code into a sandbox, where it cannot access local data or use the network. Users and applications 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.

OurGrid has been 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. The current status of OurGrid is available at http://status.ourgrid.org.

2.3. SegHidro Core Architecture

The SegHidro Core Architecture, also known as SHiCA, is a framework to support the execution of coupling models in a workflow fashion. SHiCA acts as a Grid broker in order to submit jobs to be executed on the Grid and gather their results. The Web portal is the entry point to SHiCA. When the user configures and submits a simulation, SHiCA assembles a model descriptor object for each of the cascading models. The descriptor contains all of the information necessary to describe a single model run and is used to assemble an OurGrid job and later to identify that specific run on a data catalog. Prior to executing each flow step, the system will check a catalog to see whether the result has already been computed or whether it needs to be submitted to the Grid. Note that data generated and used by SegHidro are essentially public. Thus, it is highly desirable to reuse data previously produced.

2.4. Data format

SegHidro defines minimal rules so that the models can be coupled in a workflow. Each model of the second and third levels must receive and/or return data in the specified file-based format. These formats are known by the SegHidro community as PMH (atmospheric model output) and PHR (hydrological model output). The adoption of these formats has been long debated within the SegHidro community. We have also considered using standard formats such as NetCDF [12] and HDF [13], but they appeared to be too complex and beyond our necessity. Our approach benefits from simplicity and collaboration, as one clear demand of the community was to have a format as simple as possible to facilitate the adaptation of existing models to fit in the workflow execution. A simple text format was the key here to influence the adoption of the format by engineers and scientists, who basically use Fortran to build applications.

2.5. Data management

Data reuse is a key point in SegHidro. The atmospheric models must ideally run only when absolutely necessary since their execution is too costly. This raises questions about data discovery and access. Moreover, effective data reuse is essential for knowledge sharing, which has been the main goal of SegHidro. Reusing some kinds of scientific data is not a trivial task. In many situations, the data needed by some user may be a subset of previously computed data. The problem here is how to discover that someone has already executed the desired simulation, how to find where the output result is and how to access it. Note that the data are often distributed across data producing sites.

Our approach to finding data is to use a data catalog. This is a classical solution in the data Grid research domain [14]. Currently, the SHiCA implementation accesses a single centralized catalog. The catalog is a map that associates a model descriptor object to one or more data descriptors. The data descriptors may assume different forms to allow access to distinct types of data storage. In SegHidro, two kinds of data storage are common: (1) regular binary or text files residing in the storage area of OurGrid sites and (2) OPeNDAP data servers. OPeNDAP [15] is significantly popular in the scientific community, more notably in the fields of oceanography and atmospheric sciences. The employment of OPeNDAP is mainly to encourage data sharing among institutions. In SegHidro, we use OPeNDAP in basically three scenarios: (1) to facilitate the reuse of data produced by atmospheric model runs; (2) to store and manipulate observational data collected by participating institutions; and (3) to allow the users to analyze data using their favorite OPeNDAP enabled data analysis package (Grads, Matlab, Excel, etc.).

The use of observational data is essential to validate forecasts produced by numerical models. Most regional centers do not employ a standard data format to keep the data organized. The SegHidro portal is capable of receiving observational data uploaded by users in the PMH format. The uploaded files are stored in an OPeNDAP server and become available for comparison with forecast data. The users may also run their second-level models directly with observational data as input, rather than using forecast data. The adoption of the OPeNDAP technology in this scenario improves the quality of user's results in an unprecedented way, as users can have access to a broader collection of data.

3. COLLABORATION

It can be observed that the resource distribution (human, computational and financial) among production centers, governmental agencies, universities, etc., is extremely unbalanced. The SegHidro portal intends to foster collaboration and resource sharing, as it creates a community with similar objectives. We have targeted three collaboration dimensions: sharing computational power, sharing human expertise and sharing data.

Some production centers or state governmental agencies do not have enough computational power to execute hydro-meteorological applications on their own. They are dependent on national or larger regional centers. In this sense, computational power is a key part of institutional empowerment. Using the Grid resources, SegHidro portal allows users access to high processing capabilities. The institutions are invited to donate their idle machine cycles and join the Grid. Empowerment may lead our partners to a whole new level of dependence relationship. Small centers are now able to produce, not only to consume. Another key point of dependence is the domain of knowledge. As forecasts and predictions produced by the atmospheric models are the primary input of the applications, meteorologists are one of the most important links of the chain. At the SegHidro portal, atmospheric models, previously configured or calibrated by specialists, are available for execution. The users just need to enter their input data according to their case study. Furthermore, users have access to agricultural, hydrological and hydro-geologic models from other experts, available at the portal, and can comparatively evaluate their results. The SegHidro portal allows this knowledge exchange in a common execution and communication platform.

A key challenge for this exchange is sharing data among institutions. Data produced and gathered at production centers are extremely useful and of great interest to the other members of the community. However, each center or agency has its own data format and accessing this data requires human intervention, for example, someone has to mail CDs or to leave the data on a FTP server. SegHidro has established a very simple text-based format to allow data to flow on its model cascade, and that is also being used to exchange data among institutions. Data results produced by the execution of SegHidro portal applications may also be shared. The reuse of results is extremely positive since producing them may be very time consuming and requires someone else's expertise to configure the model parameters.

4. CASE STUDIES

In order to better understand the applicability of the SegHidro portal, let us evaluate two examples of how SegHidro benefits the water management community. The first example concerns applying SegHidro to mitigate a water scarcity problem of a region. The second example concerns a region that suffers with flooding.

4.1. Reservoir management

Campina Grande is a major urban center in the semiarid and drought-prone north-eastern Brazil. Approximately 500 000 inhabitants are supplied by the Epitácio Pessoa reservoir, which also provides water for irrigation and fisheries. The reservoir contributing basin lies in the driest region of Brazil, where rainfall is concentrated in only four months and presents a high inter-annual variability.

DOI: 10.1002/cpe

The last decade was particularly dry, with two of the most severe drought periods of the century, around the years 1993 and 1998. By the end of 1999, the reservoir experienced its lowest storage level and worst water quality indices since it was built in the 1950s. Irrigation was almost stopped and, during several months, only 46% of the usual discharge for urban consumption could be delivered, with severe social and economic consequences. In November 1999 the reservoir storage was the lowest in its history (about 15% of its capacity) and the salt concentration reached high levels, so that the water was no longer drinkable. Fortunately, the rainy seasons of years 2000, 2004 and 2005 were very wet, and restored the reservoir storage to its full capacity and water quality to a reasonable level. From year 2000 onwards, rainfall, runoff and reservoir storage forecasting information could be produced by the local agency through model integration. However, they could not produce rainfall forecasts, being dependent on information from larger forecast centers. Due to limitations in data assimilation from the national forecast center, the process used to be conducted only for a few scenarios and, many times, using only aggregated probabilistic rainfall forecasts instead of the actual numerical forecasts.

Using the SegHidro portal, the agency can now run regional atmospheric models to generate rainfall forecasts. Flow discharges into the reservoir are estimated by hydrological models, and are input to reservoir management models, which use present information on the reservoir's storage and estimates of future lake evaporation and consumers' water demands. These can be as simple as a simulation model for forecasting future storage given a certain water-demand scenario, or complex optimization and knowledge-based decision models. The multiple scenarios generated, when combining ensemble rainfall forecasts with water-demand estimates and decision alternatives, are vital to allow better sensitivity and uncertainty analysis throughout the water management decision process. Using the portal, this process was simplified and the quality of the forecasts will hopefully be improved. This experience demonstrates our hypothesis that the Grid can provide much more than high-performance computing. By means of the portal, integrated and, therefore, more useful and appropriate management information is being produced by a network of researchers and operational centers of distinct disciplines.

4.2. Flood alerts

In early June 2005, an easterly wave hit the Brazilian northeast causing flooding in several coastal cities. The metropolitan area of Recife, one of the largest cities of the region, was among the most affected areas. In such a situation, the civil defense demands information about weather forecast to be as precise as possible. Figure 3 shows the enhancement of a forecast through downscaling.

In the first model, with a horizontal resolution of about 78 km, one cannot distinguish between the east coast that was under the rainy system and the semi-arid portion of Pernambuco State, which was already in its dry season. The first grid of downscaling, with a resolution of 20 km, allows one to see the separation between the coastal system and the semi-arid area. However it is the second downscaling grid, with a 5 km resolution, that allows the civil defense personnel to see in detail the spatial distribution of forecasted rain. The temporal resolution of this last model is 15 minutes. Using the SegHidro portal, it is possible to reuse such information, whose generation demands high processing power, and to run a hydrological model to forecast the areas of higher risk and to improve civil defense actions.

DOI: 10.1002/cpe

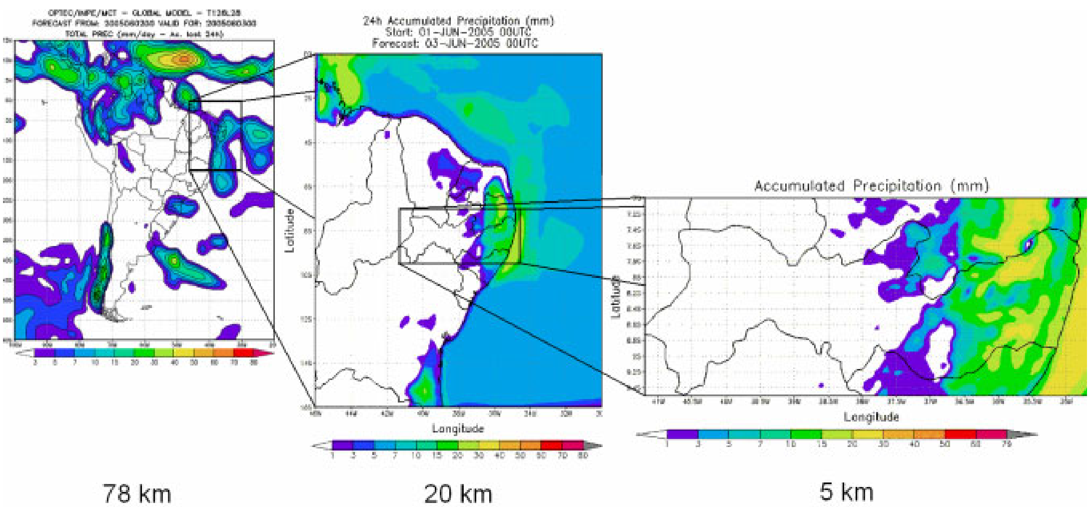

Figure 3. A cascade of the model from the global model, with a 78 km resolution, to a regional model, with a 5 km grid mesh. The first highlighted area is northeast Brazil and the second is the state of Pernambuco.

5. RELATED WORK

There are other initiatives that, similarly to SegHidro, intend to provide an infrastructure to environmental research, such as LEAD [16] and ESG [16]. 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.

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 perform meteorology research and education. They focus on numerical weather prediction based on meso-scale models. Similarly to SegHidro, LEAD models the execution of the applications as workflows and has strong components of data management. Also, a Web portal is the system front-end.

The ESG project targeted their efforts towards the problems faced by the climate modeling community. In order to execute their models, high computing power and an archival area able to store hundreds of gigabytes or some terabytes of output data are necessary. Furthermore, in the ESG context, for the user to consume these data it is necessary to move them to the user's desktop. The ESG prototype development was based on some data Grid technologies [16], such as replica management, data transfer, request management and so on. Data transfer is a central issue on ESG. It employs GridFTP to cope with this issue. The SegHidro system also needs to deal with this problem; however, the approach is to diminish data transfers by moving the computation to the data.

DOI: 10.1002/cpe

6. CONCLUDING REMARKS

In this paper we have presented our experience in building the SegHidro portal and have analyzed how it fosters collaboration among hydrologists, meteorologists and decision makers who have a common interest, namely the water issue in the Brazilian northeastern region. SegHidro leverages collaboration in many aspects: computational power, data and human expertise. In particular, SegHidro allows people with complementary expertise to combine their talents together into a water allocation analysis. A SegHidro application is a workflow of scientific simulation models, where one model's output serves as input to its succeeding model, which is called model coupling. The system makes it easier for researchers to 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 that 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 and operational meteorologists, hydrologists and managers to effectively use it. A key part of such simplicity is due to choosing a very simple text-based data format that allows the model coupling. Moreover, we recognize 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 perform large parameter sweeps.

In short, the main contribution of SegHidro was to create an infrastructure that fosters collaboration among researchers and decision makers: meteorologists produce 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 easier to evaluate how one's 'piece of the whole' contributes to the final result, thus providing invaluable feedback for the researchers to refine their models.

ACKNOWLEDGEMENTS

Authors are grateful to the SegHidro team and to the project partners, particularly the Center for Weather Forecasting and Climate Studies (CPTEC/INPE) and Ceará Foundation for Meteorology and Water Resources (FUNCEME). The SegHidro project is sponsored by the Brazilian Ministry of Science and Technology (FINEP/CT-INFO) under grant 01.04.08888.00. Walfredo Cirne would like to acknowledge the financial support of CNPq/Brazil (grant 302.317/03).

REFERENCES

1. Foster I, Kesselman C. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann: San Mateo, CA, 1998.
2. Dahan M, Mueller K, Mock S, Mills C, Regno R, Thomas M. Application portals: Practice and experience. *Concurrency and Computation: Practice and Experience* 2002; **10**(13–15):1427–1443. DOI: 10.1002/cpe.694.
3. Fayad M, Schmidt D, Johnson R (eds.). *Building Application Frameworks: Object-Oriented Foundations of Framework Design*. Wiley: New York, 1999.

4. Gannon D, Fox G, Pierce M, Plale B, von Laszewski G, Severance C, Hardin J, Alameda J, Thomas M, Boisseau J. *Grid Portals: A Scientist's Access Point for Grid Services*. Global Grid Forum: Lemont, IL, 2003. Available at: http://forge.gridforum.org/projects/ggf-editor/document/GCE-Portal-working-draft/en/1/GCE-Portal-working-draft.pdf.
5. Wang Y, Leung LR, McGregor JL, Lee DL, Wang WC, Ding W, Kimura F. Regional climate modeling: Progress, challenges and prospects. *Journal of the Meteorological Society of Japan* 2004; **82**(6):1599–1628. DOI: 10.2151/jmsj.82.1599.
6. Jetspeed. http://portals.apache.org/jetspeed-1 [September 2005].
7. Cirne W, Brasileiro F, Andrade N, Santos R, Andrade A, Novaes R, Mowbray M. Labs of the World, unite!!!, *Technical Report*, Universidade Federal de Campina Grande, Departamento de Sistemas e Computação, 2006 (*Journal of Grid Computing*, accepted for publication).
8. Andrade N, Brasileiro F, Cirne W, Mowbray M. Discouraging free-riding in a peer-to-peer CPU-sharing Grid. *Proceedings of the 13th IEEE International Symposium on High-Performance Distributed Computing (HPDC13)*, June 2004. IEEE Computer Society Press: Los Alamitos, CA, 2004; 129–137. DOI: 10.1109/HPDC.2004.9.
9. Santos-Neto E, Cirne W, Brasileiro F, Lima A. Exploiting replication and data reuse to efficiently schedule data-intensive applications on Grids. *Proceedings of the 10th Workshop on Job Scheduling Strategies for Parallel Processing* (*Lecture Notes in Computer Science*, vol. 3277), June 2004, Feitelson D, Rudolph L, Schwiegelshohn U (eds.). Springer: Berlin, 2005; 210–232. DOI: 10.1007/11407522_12.
10. Netto MAS, de Rose CAF. CRONO: A configurable and easy to maintain resource manager optimized for small and mid-size GNU/Linux Clusters. *Proceedings of the 2003 International Conference on Parallel Processing (ICPP'03)*, October 2003. IEEE Computer Society Press: Los Alamitos, CA, 2003; 555–562. DOI: 10.1109/ICPP.2003.1240623.
11. Barham P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, Neugebauery R, Pratt I, War A. Xen and the art of virtualization. *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOPS'2003)*. ACM Press: New York, 2003; 164–177. DOI: 10.1145/945445.945462.
12. NetCDF. http://www.unidata.ucar.edu/software/netcdf [September 2005].
13. National Centre for Supercomputing Applications. HDF5—a new generation of HDF. http://hdf.ncsa.uiuc.edu/HDF5 [August 2005].
14. Chervenak A, Foster I, Kesselman C, Salisbury C, Tuecke S. The data Grid: Towards an architecture for the distributed management and analysis of large scientific datasets. *Journal of Network and Computer Applications* 2000; **23**(3):187–200. DOI: 10.1006/jnca.2000.0110.
15. Cornillon P, Gallagher J, Sgouros T. OPeNDAP: Accessing data in a distributed, heterogeneous environment. *Data Science Journal* 2003; **2**:164–174.
16. Allcock B, Foster I, Nefedova V, Chevrenak A, Deelman E, Kesselman C. High-performance remote access to climate simulation data: A challenge problem for data Grid technologies. *Proceedings Supercomputing 2001*. Elsevier: Amsterdam, 2001; 20. DOI: 10.1109/SC.2001.10020.
17. Droegemeier K, Chandrasekar V, Richard C. Linked Environments for Atmospheric Discovery (LEAD): Architecture, technology roadmap and deployment strategy. *Proceedings of the 21st Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*. American Meteorology Society: San Diego, CA, 2005.