

AdriaScience Gateway: Application Specific Gateway for Advanced Meteorological Predictions on Croatian Distributed Computing Infrastructures

D. Davidović*, T. Lipić* and K. Skala*

* Ruđer Bošković Institute/Centre for Informatics and Computing, Zagreb, Croatia
E-mail: {ddavid,tlipic,skala}@irb.hr

Abstract - Marine traffic and rapidly growing tourism in the Adriatic Sea region are the main reasons for investigating different weather phenomena and developing the prediction models that depend on distributed computing infrastructures. Weather phenomena like storms and waterspouts are common in the in the Mediterranean Basin and thus in the Adriatic Sea region. Their occurrence is becoming easier to follow in past few years thanks to the advanced numerical weather prediction models and modern communication and computing technologies. This paper presents AdriaScience gateway, the application specific gateway for meteorological applications based on WS-PGRADE/gUSE technology developed within the SCI-BUS project. The main goal of the application specific gateway is to facilitate a universal access to the different Croatian distributed computing infrastructures for Croatian meteorology community.

I. INTRODUCTION

The weather conditions have a great impact on the economy especially on tourism and marine traffic that are rapidly growing in the Adriatic Sea region in the past few years. Severe weather phenomena like storms and waterspouts are common in the in the Mediterranean Basin and thus in the Adriatic Sea region. Storms and waterspouts attract attention not only because of their unique appearance but also because of their ability to produce high risk to the community and also significant property damages. Thus, investigation of this kind of weather phenomena and accurate and forehead prediction of their occurrences is important for Croatian economy and society.

Different techniques and models are used to predict weather conditions and special phenomena that occur in the Adriatic Sea region. These techniques depend on gathering the meteorological data from the observatory stations and weather radars or using advanced meteorological prediction models from atmospheric sciences.

The atmospheric sciences of today are advancing towards predicting very complex phenomena at ever

smaller temporal and spatial scales. These phenomena have large impacts on both scientific research and operational weather forecast. One of the principal tools utilized in these efforts are numerical weather prediction models. These numerical models usually demand large execution time and resource allocations. Therefore, large parallel systems are required in order to execute them. One of such numerical prediction models is public domain WRF-ARW model (Weather Research and Forecasting model - Advanced Research WRF) [1]. There are some non-free models like ALADIN [2] model that are mostly used by the national weather and forecasting centers.

As numerical weather prediction models require very large computational and storage resources, most of them are developed to exploit different distributed computing infrastructures (DCI) like Clouds, Grids and High Performance Computing (HPC) systems. Accordingly, many of them have been extensively developed and adopted for different DCIs in the last decades. For example, the WRF-ARW model is developed for the cluster and HPC systems but some research has also been done in porting the model on the Grid: SEE-GRID infrastructure [3] and EGEE/EGI infrastructure [4]. The drawback is that the process of the adaptation for each application is not unique and it requires, for the end-user, a certain level of technical knowledge of each infrastructure in order to be properly used.

Therefore, the ultimate goal of the future development would be to make a transparent and uniform access to different DCIs which will hide from the end user the architecture-specific implementation details. One solution is the development of scientific gateways that will hide the infrastructure API and allow users to execute their applications without even knowing on which infrastructure their application is executed. In this paper we will describe how we developed the AdriaScience application specific gateway that utilizes Croatian distributed computing infrastructures for meteorological applications. The gateway is based on the WS-PGRADE/gUSE [5] generic purpose gateway developed within the EU FP7 SCI-BUS project [6].

II. RELATED WORK

In the last few years there were numerous adaptations of weather prediction models on different grid infrastructures. The development of AdriaScience application specific gateway heavily depends on our previous implementation of the WRF-ARW model on the SEE-GRID infrastructure [7] and WS-PGRADE/gUSE framework for easier development of custom Scientific Gateways for various distributed computing infrastructures.

A. WRF-ARW for SEE-GRID

The WRF-ARW model framework on the grid infrastructure is a collection of the user-specific command line execution scripts that are executed on the front-end. The scripts are responsible for submitting model to the grid Workload Management System (WMS), allocating the grid computing elements (CE) and retrieving data from the grid storage elements to the front-end or predefined user machine. The terrestrial data and initial/boundary conditions data are automatically downloaded from the external servers or copied from the grid storage elements. Upon successful download, the WRF model is started. The grid workflow starts with the execution of the WRF pre-processing system (WPS) on one grid computing element, after which the core application is started and the output data are stored on predefined grid storage elements. Users can download model outputs from storage elements using the LCG-tools, DM-Web [8] application service developed within SEEGRID project or using customized scripts for automatic WRF model's data collecting.

Submission of a job is the process that sets all model parameters, describes the job using JDL (Job Description Language), updates the WRF model description files and submits model to the grid. Regarding the job described, WMS allocates a grid CE where the user's job will be executed. The main execution script is responsible for defining and setting almost all parameters needed for the model execution. Based on the user command line parameters the script sets model execution type (scientific or operational), number of processors, and model input files and other specific parameters.

As described above, the process of starting the applications and retrieving the data is complicated for the end user who does not have, at least, basic technical knowledge of the grid systems. The process of preparing and submitting the model to the grid is iterative and require series of steps that have to be initialized by the user. Furthermore, the implementation is very specific to the SEE-GRID technologies and it not directly scalable to other DCIs but need a certain adoptions and changes in the execution scripts.

By using Scientific Gateways it is possible to override the scalability and usability problems of the traditional implementation approaches. Accordingly, we have decided to use WS-PGRADE/gUSE framework for the development of AdriaScience application specific gateway that will enable process of customizing, executing and monitoring WRF model workflows easier.

B. Scientific Gateways

Scientific Gateways are frameworks (or sets of tools) which integrate applications, data and tools to enable running applications on distributed computing infrastructures. They are usually integrated via a portal or a suite of applications that can be further customized to meet the needs of a specific community. The gateways enable entire communities of users associated with a common discipline to use computing and storage resources from different infrastructures (national or private) through a common user-friendly interface. In this way, users can focus on their applications instead of learning and managing the complex underlying infrastructure.

Scientific Gateways are becoming to play an important role in many scientific research areas and are being developed for different national and private infrastructures. It is worth mentioning the Scientific Gateways developed on United States infrastructures: Extreme Science and Engineering Discovery Environment (successor of TeraGrid; XSEDE) infrastructure and National Energy Research Scientific Computing Center (NERSC) infrastructure and European Grid Infrastructure (EGI) Scientific Gateways.

Several Scientific Gateways for meteorological and global climate research communities were also emerged in United States and Europe. In United States, there were developed NASA Earth Science Gateway[9], LEAD portal[10] and two Earth System Grid Gateways: one for XSEDE[11] infrastructure and other for NERSC[12] infrastructure. In Europe for EGI infrastructure exists Web-based EarthServer Science Gateway [13].

The relevance of Science Gateways will further increase with the development of more sophisticated user interfaces and easier access mechanisms for many different distributed computing infrastructures. WS-PGRADE/gUSE is a framework that enables easier development of user-specific gateways for various distributed computing infrastructures.

C. WS-PGRADE/gUSE

WS-PGRADE/gUSE is an open source framework for development of generic workflow-based DCI gateways which is being developed within the SCI-BUS project [6] that allows customization and development of the application-specific gateways according to users' needs. The generic gateway supports submission of the applications on different DCIs including clusters (LSF, PBS), service grids (Globus, gLite, ARC, Unicore), BOINC desktop grids, and clouds (Google App Engine, CloudBroker).

The main part of the gateway is gUSE (grid and cloud User Support Environment). gUSE is a virtualization environment that provides a scalable set of high-level services for managing different DCIs. Each DCI is handled by its specific plugin that can be easily added allowing portal developers to add support for their specific DCI.

WS-PGRADE portal is a web-based user interface for gUse services. It is based on Liferay [14] portal

technologies. WS-PGRADE uses the client APIs of gUSE services to turn user requests into sequences of gUSE specific Web service calls. A generic portal can be easily extended to the application specific portal via the Application Specific Module (ASM) [5]. ASM is an API web interface which provides a set of tools for developing customized portlets.

The main feature of the WS-PGRADE/gUSE portal is the workflow support. The portal incorporates a set of tools and editors that enables creation, management and storage of the application workflows. Furthermore, the portal also provides a support for embedded workflows, i.e. workflows that are part of other workflows. In addition the workflows can be shared among users directly on the portal or via the SHIWA Simulation Platform [15]. The portal also features the simultaneous using of multiple DCI for the same workflow. Therefore, it is possible to off-load parts of the workflows on the computing infrastructures that are the most suitable for the particular parts of the application workflow.

III. ADRIASCIENCE GATEWAY

The purpose of WS-PGRADE/gUSE based AdriaScience Scientific Gateway is to provide easier usage of meteorological applications on Croatian distributed computing infrastructures. Currently, the focus is put on integration of WRF-ARW model which will later enable easier integration of other meteorological applications such as ALADIN or ROMS. The WRF-ARW model is integrated as scientific workflow that can be used through user-friendly portlets enabling robust specific user scenarios.

A. Supported DCIs

The AdriaScience gateway supports two national distributed computing infrastructures both operated by the University Computing Centre (SRCE), University of Zagreb.

Croatian National Grid Infrastructure (CRO-NGI) [16] is the largest distributed computing infrastructure in Croatia. It consists of 5 grid sites with more than 1180 computing cores, 36 graphical processing units (GPUs), and 47TB of the storage capacities. CRO-NGI sites contain basic services of Globus Toolkit 5 (GT5) grid middleware system. Accordingly, we use GT5 DCI

Bridge Plugin from gUSE for access to this infrastructure.

Isabella computing cluster [17] is the largest Croatian HPC system with 480 CPU cores and 2 TB of the main memory. Also it provides 36TB of the shared storage space. On the Isabella are more than 180 registered user and more than 58 projects. Isabella uses Son of Grid Engine (SGE) job management system. To access Isabella resources we developed gUSE DCI Bride Plugin for SGE.

B. Scientific workflows

One of the most important parts of the WS-PGRADE/gUSE portal are scientific workflows. A workflow is defined as a sequence of the connected steps (logical parts) that describe the overall application execution throughout time. A workflow also describes data flow between the individual parts and its dependencies. Our objective is to create workflows that will describe our specific meteorological problems with the emphasis on the WRF-ARW prediction model. The workflows are uniformly describing the data flow and the certain parts of the WRF-ARW execution process.

WRF-ARW model is integrated as a gUSE scientific workflow composed of 7 processes. The composition is structure of directed acyclic graph defined using WS-PGRADE/gUSE Graph Editor tool. Fig. 1 depicts the composition of WRF-ARW workflow. The workflow is described generally because it should be applicable of various DCIs. Each yellow square represent one individual computational part of the WRF-ARW model, green boxes input files/data, grey output data and the lines between them the data flow.

WRF-ARW workflow processes are:

- **IBC-data – data collection.** In this process are downloaded initial/boundary condition data from remote servers and stored on the DCI's storage elements. The path to the condition data location on storage elements is passed as an input to *ungrib* process.
- **geogrid – pre-processor.** Define WRF grid, generate map, elevation and land information. As an input requires the location of the terrestrial data. The recommendation is to execute *geogrid* process on the same DCI computing element as *ungrib* and *metgrid*.

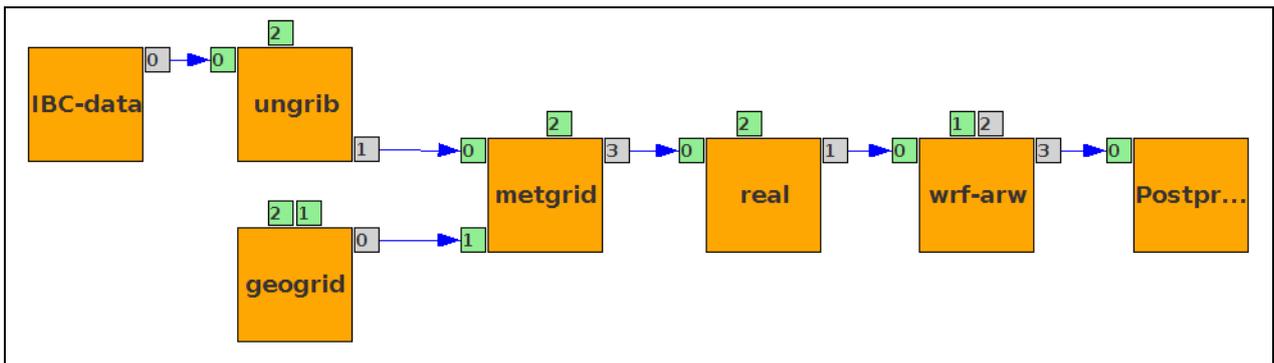


Fig. 1 WRF-ARW WS-PGRADE/gUSE workflow

- **ungrib – pre-processor.** This process unpacks and prepares real-data analyses/forecasts from another (global) models. The inputs are real-data previously stored on the targeted DCI's (CRO-NGI or Isabella) storage system in the *IBC-data* process.
- **metgrid – pre-processor.** This process interpolates the data provided from *ungrib* and *geogrid* processes to the WRF grid.
- **real – real-data initialization.** The process initializes real-date and generates input and boundary files for the core program. It requires inputs from *metgrid* process.
- **wrf-arw – core WRF-ARW program.** This is the main WRF-ARW program. The program requires parallel environment (MPI or OpenMP). It is executed on one or more DCI computing elements depending on the configuration. Requires data from *real* process.
- **post-processing –** This process includes embedded workflows for post-processing. Different workflows can be added here depending on the required post-processing operation. It may include visualization, data extraction or data control and verification. Inputs are data in NetCDF file format generated in *wrf-arw* process.

C. AdriaScience portlets

The main end-user tools of the AdriaScience gateway are the application specific portlets. The portlets are used to make configuration of the application workflows easier and faster. The new AdriaScience portlets will extend the functionality mostly in terms of remote data management and configuration of pre-defined specific application workflows. Also, some existing generic WS-PGRADE/gUSE portlets (e.g. authentication and resource configuration) will be used as well.

The WRF-ARW specific AdriaScience portlets are:

- Portlet for managing data on Global Storage System (GSS -<http://code.google.com/p/gss/>) and Gfarm. The portlet will enable management of users' data on remote storage systems. This portlet will be integrated in the WS-PGRADE Storage menu.
- Pre-processing portlet for preparing static (terrestrial) and dynamical (boundary condition) data. Using this portlet the users will be able to define the start and the end date of the simulation, upload configuration files, the path on the storage systems (Gfarm or GSS) where the data are stored, and the computing resource. The portlet will be integrated in the End User menu.
- Application specific portlet for each pilot application (e.g. WRF-ARW). The portlets will allow users to define the specific workflow

parameters, upload the model input data and upload application configuration files (e.g. *namelist.input* file for WRF-ARW). The application submission portlets will be integrated in the End User menu.

- Post-processing portlet. This portlet will allow users to define post-processing visualization tools. The functionality of the portlet is to define the path of the files to be processed (as returned by the models and stored on GSS or Gfarm) and the specific parameters for visualization tools.

D. Specific user scenarios

Specific user scenarios describe potential usages of the WRF-ARW model workflows and on the Croatian distributed computing infrastructure. The steps of the use-case scenario are:

- User authentication on the DCI (CRO-NGI or Isabella cluster). The authentication is done in the "Security" portlet by retrieving user certificate.
- Preparation of the terrestrial and initial and boundary condition data. The data are downloaded from local computers or remote servers to the publicly available storage systems (national resource operated by SRCE – Gfarm and GSS)
- Configuration of the workflow. Setting the preferred DCI site, number of processors, WRF-ARW description files (*namelist.input* and *namelist.wps*), and defining the remote storage resources for intermediate and final output files.
- Submitting the workflow and monitoring the execution process and intermediate files.
- Remote data management of the intermediate and final output files.
- Post-processing of the model output data including visualization and data management

IV. CONCLUSION

The main purpose of AdriaScience Scientific Gateway is to provide easier and intuitive usage of meteorological applications for meteorological and global climate research communities on Croatian distributed computing infrastructures. We focused on integration of WRF-ARW model as workflow and application specific portlets using WS-PGRADE/gUSE framework. While AdriaScience Scientific Gateway facilitates the conduction of robust user scenarios, WS-PGRADE/gUSE framework facilitates the development of such user and application specific workflow-based DCI Scientific Gateways.

For future work we plan extending AdriaScience Scientific Gateway with additional meteorological models such as ALADIN, ROMS or Source-sink model (SSM). This will additionally enable creation of nested workflows that combines two or more models for different complex use cases giving impetus to the formation of larger user community. In addition to extending set of meteorological applications we will also consider ways how to provide

different application specific data management functionalities using application specific portlets.

ACKNOWLEDGMENT

This work is partially funded by the EU FP7 project SCI-BUS.

REFERENCES

- [1] W. Wei, C. Bruyère, M. Duda, J. Dudhia, D. Gill, H. C. Lin, J. Michaelakes, S. Rizvi, X. Zhang, "WRF-ARW Version 3 Modeling System User's Guide," Mesoscale&Microscale Meteorology Division, National Center for Atmospheric Research, Boulder, USA, 2009.
- [2] ALADIN: High Resolution Numerical Weather Prediction Project, <http://www.cnrm.meteo.fr/aladin/>, 2013.
- [3] D. Davidović, K. Skala, D. Belušić, M. Telišman-Prtenjak, "Grid implementation of the Weather Research and Forecasting model," Earth Science Informatics, vol. 3, pp. 199-208, 2010.
- [4] V. Fernández-Quiruelas, L. Fita, J. Fernández, A. S. Cofiño, J.M. Gutiérrez, "WRF4G: simplifying atmospheric modeling experiment design in distributed environments," IBERGRID conference, pp. 327-332, 2011.
- [5] P. Kacsuk, Z. Farkas, M. Kozlovsky, G. Hermann, A. Balasko, K. Karoczkai, and I. Marton, "WS-PGRADE/gUSE Generic DCI Gateway Framework for a Large Variety of User Communities," Journal of Grid Computing, vol. 10, pp. 601-630, 2012.
- [6] SCI-BUS home page, <http://www.sci-bus.eu/>, 2013.
- [7] D. Davidović, K. Skala, "Implementation of the WRF-ARW prognostic model on the Grid," Proceedings MEET&GVS 33rd International Convention MIPRO 2010, Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO, vol. 1, pp. 253-258, 2010.
- [8] B. Marović, M. Potočnik, D. Okiljević, J. Oklobdžija, "DM-Web – Web access and management of application-specific files in LFC catalogue," Proc. of the SEE-GRID-SCI N. Alameh, M. Bambacus, J. Evans, S. Marley, "NASA's Earth Science Gateway: A Platform for Interoperable Services in Support of the GEOSS Vision," IGARSS 2006.
- [9] N. Alameh, M. Bambacus, J. Evans, S. Marley, "NASA's Earth Science Gateway: A Platform for Interoperable Services in Support of the GEOSS Vision," IGARSS 2006.
- [10] Christie, M., & Marru, S. (2007). The LEAD Portal: a TeraGrid gateway and application service architecture. Concurrency and Computation: Practice and Experience, 19(6), 767-781.
- [11] ESG-NCAR Gateway, <http://www.earthsystemgrid.org>
- [12] ESG - Earth System Grid Climate Gateway and Data-node, <http://esg.nersc.gov/>
- [13] EarthServer Science Gateway, <http://earthserver-sg.consortio-cometa.it/>
- [14] Liferay portal homepage, <http://www.liferay.com/>, 2013
- [15] SHIWA – Sharing Interoperable Workflows for large-scale scientific simulation on Available DCIs. <http://www.shiwa-workflow.eu>
- [16] Croatian National Grid Infrastructure – CRO-NGI, <http://www.cro-ngi.hr>, 2012.
- [17] Isabella cluster, <http://www.srce.unizg.hr/homepage/products-services/computer-resources/cluster-and-grid-technologies/isabella-cluster/>, University of Zagreb, University Computing Centre