Design of Grid Computing Environment for Operative Weather Nowcasting

Giovanni Aloisio[††], Dario Conte[†, ††], Gian Paolo Marra[†, ††], Franco Prodi[†] and Gianvito Quarta[†, ††]
†) Institute of Atmospheric Sciences and Climate of the Italian National Research Council
via per Monteroni, 73100 Lecce – Italy
††) Center for Advanced Computational Technologies/ISUFI, University of Lecce,
at NNL/INFM&CNR Lecce, Italy
{d.conte, gp.marra, f.prodi, g.quarta}@ isac.cnr.it, {giovanni.aloisio, dario.conte, gianpaolo.marra,
gianvito.quarta}@unile.it

Abstract - *Weather nowcasting is a meteorological forecast which alerts civil protection institutions in case of heavy rain events, tornadoes, fires or heat waves and so on. This technique needs to process different data in a very short time. This paper, critically describes the design process of a computational infrastructure for weather nowcasting using both web and Grid technologies as a suitable solution for the use of nowcasting in an operative and concrete context.*

Keywords: Grid Computing Environment, Weather Nowcasting.

1 Introduction

Weather forecasting has become more and more important in recent years. Short-range weather forecasting has improved significantly and is nowadays applied in several activities: maritime transport, civil and military aviation, determination of energy requirements for human activities and planning of human activities. An accurate knowledge of the evolution of weather conditions is really useful in order to alert civil protection institutions in case of heavy rain, tornadoes, fires or heat waves.

Weather nowcasting [1] represents a meteorological forecast from 0 to 12 hours, founded on the information already present. Numerical models of forecasts operationally used by the meteorological services cannot foresee the short term weather forecast within the necessary resolution of space-timing, inside this temporal interval. Different procedures type based on observations of the current weather conditions (observations on the ground, from radar and satellite) are required to get detailed information on short term weather and eventually alert it in timely manner.

Thanks to meteorological radar and the ability to foresee fields of precipitation with elevated resolution has always been considered a valid aid for short term forecasts. Moreover, geostationary satellites such as the Meteosat 8 of EUMETSAT (launched in 2002 and first of the Meteosats Second Generation family), provides the possibility to observe the atmosphere and cloud coverage, from the synoptic scale up to the cloud coverage scale, every 15 minutes. The Meteosat 8 clouds and precipitation products for weather nowcasting and the short-term meteorological forecast have been continually refined in recent times, they are surely abandoning the context of the purely scientific applications and they have reached applicability to the operational room to the forecaster's console.

At the same time, meteorological models represent the fundamental tool to study the atmospheric dynamics and its evolution. The large number of variables involved, the feedback processes and the non linear interactions among systems of different scale which make the problem extremely complex and therefore make the use of the model tool necessary to correctly represent the complexity of the atmospheric motions. In fact, the numerical models have recently acquired more elaborate physical parameterizations that together with the availability of high performance computers, have allowed them to meaningfully increase the accuracy of forecasts which can be available in a brief time period. Grid computing [2, 3] today provides a favorable and consolidated way to combine, access and exploit computational power and this can radically change the traditional meteorological models run concept.

In this paper we describe the design of a computing infrastructure able to answer, with effectiveness and performance, the imposed requirements for weather nowcasting, allowing the synergic integration of the above mentioned kind of data. The proposed solution allows the disbursement of high-level services to the experts of the application domain through an efficient and coordinated use of the necessary computational resources and through the heterogeneous data sources virtualization.

The implementation of the infrastructure will be based on the Grid computing as motivated, which is the best technology that allows the efficient and coordinated use of heterogeneous resources geographically distributed. Moreover, Grid approach to the problem guarantees to achieve the needed level of abstraction from aspects such as data sources and computational resources.

This paper is organized as follows. Section 2 recalls the weather nowcasting definition, data integration and processing. Section 3 describes the design of the infrastructure and in section 4 a discussion about the design pattern and the implementative choices are given. Finally, section 5, concludes the paper and describes future directions.

2 Data Integration and Processing in Weather Nowcasting

In this section, we describe the weather nowcasting process and its requirements with respect to user and system points of view.

2.1 Radar-meteorology

Radar meteorology is based on a complex and crucial physical process: the interaction between electromagnetic waves and cloud droplets and crystals, snow flakes and rain drops. While satellites allow the macroscopic evaluation of many atmospheric parameters, the meteorological radar provides detailed information about the composition of clouds and precipitation. Consequently the meteorological radar is a fundamental instrument both for the research in cloud physics and for the operational very short term forecasting (nowcasting).

The availability of Doppler measurements adds the interesting possibility to evaluate the radial component of the velocity of the meteorological target.

2.2 Meteorology from satellite

The geostationary satellites, such as Meteosat 8, have the unique possibility to observe the atmosphere and its cloud coverage from the synoptic scale up to a cloud coverage scale every 15 minutes. Therefore, their sensors constitute the essential tool for the nowcasting observations and meteorological forecasts from 0 to 12 hours. The observations from the satellites in polar orbit complete the observations from space. The availability of an increasing number of spectral bands in the visible, near infrared, infrared and microwaves with spatial high resolution and short reexamination time has meaningfully contributed to the planning of new products that are now in an advanced phase of testing and, in some cases, are already operationally used for weather monitoring and the weather forecast.

2.3 Non meteorological satellites

The acquisition of remote sensing data from satellite, together with punctual type data on the ground, is fundamental for nowcasting. Particularly the satellite data has a double action: as a primary informative source for the numerical model in use and as a source of territorial information to be used in the refinement phase of the models and adaptation to the specific local environmental conditions. In this context the possibilities offered by the orbiting sensors are particularly useful to get both maps of classification of the cloud coverage and measure the dampness on the ground, the key parameter for the study of the environmental and meteorological phenomena.

2.4 Meteorological modelling

Meteorological modelling is the fundamental tool to study the atmospheric dynamics and its evolution. Recently, numerical models have been updated with more elaborate physical parameterizations that allow to significantly improve the accuracy of the forecasts and their availability in a short period of time. This is a necessary requirement for an operative use of the system. Obviously the accuracy of models and the big number of variables involved requires the increase of supplied computing power, already elevated.

2.5 Requirements analysis

Considering the above mentioned process and requirements for weather nowcasting it is extremely important to use an appropriate computing platform (such as high performance resources), able to reduce the computing times and thus to allow real time (or near real time) usage.

Moreover, the heterogeneity of involved data (radar, meteorological and non meteorological satellite data) imposes to introduce a homogenization service, able to automatically convert the input provided by different sources in terms of geographic reference system, resolution, data format and so on.

The same requirement can be considered from a users point of view: in fact, the computing infrastructure must allows them to collect data coming from involved data sources and it must provides the environment for composition, integration and execution of forecast models.

Taking into account all these faces, we need to design a computing infrastructure able to allow and to collect the data from the available sources, able to supply the environment for the integration and the execution of the prognostic models and finally, able to publish the output on different communication channels.

Moreover, in order to allow an operative usage of methodology, the system must provide services for reliable prognostic models execution such as automatic and/or the on-demand elaboration of the algorithms. Such services must allow to activate the computation on appropriate hardware platforms and to guarantee an appropriate level of performance (for instance, through a performance contract) and fault tolerance. The services must expose an appropriate interface that can be specialised in the applicative domain, thus supplying a set of high level functionalities. Also M2M (Machine-to-Machine) services must be provided to allow the access to the forecasts, supplied to other information systems, too.

3 Computing Environment Design

The computing environment will provide all functionalities to collect different data sources for harmonization data, forecast models execution and results publishing. This data normalization is oriented to data transposition according to a common geographical reference system and a common intermediate processing level. The services oriented to execute forecasting models will be such that processing can be automatic and/or on demand based on designed algorithms. These services are such that processing will run on appropriate hardware platforms with high performances and fault tolerance features. So, designated architecture will be implemented using a modular strategy, in order to ensure an easy functionality extension as well as scalability for growing resource prospect. In particular we can identify three main levels (as depicted in Figure 1) they are:

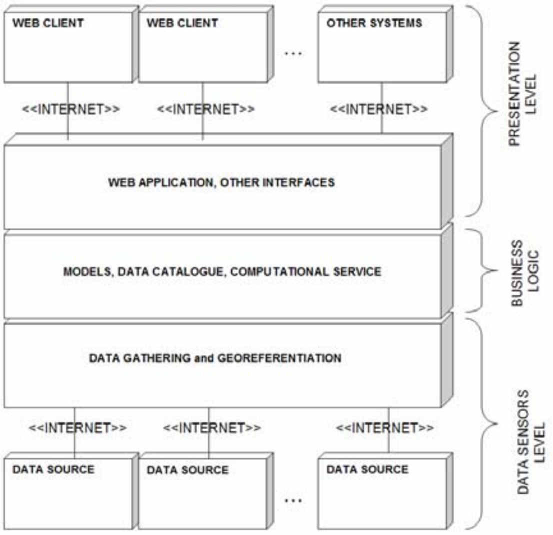

Figure 1. Overview of the proposed architecture.

i. **Data Sensors:** in this level information coming from different data sensors in their native format belonging to underlying acquisition system, will be normalized and stored in a common format chosen for the system;

ii. **Business Logic:** acts as software modules set that execute acquired data processing. This level will include all forecast model components, data-fusion algorithms and service modules. Service modules will manage execution of each algorithm on appropriate computing platforms. The grid services will be positioned at this level which will allow the use the external computational resources in a transparent manner;

iii. **Presentation Level:** will include all the components that implement the final users and systems' interfaces.

3.1 Data Sensors

At this level, raw information will be collected from various sources and transformed into a common format. This level will provide data publication services which maybe accessed by legacy systems. In fact it will have a unique front-end for heterogeneous sources access and will provide a standard interface with guaranteed efficiency, security, transparency, modularity, extensibility and portability features.

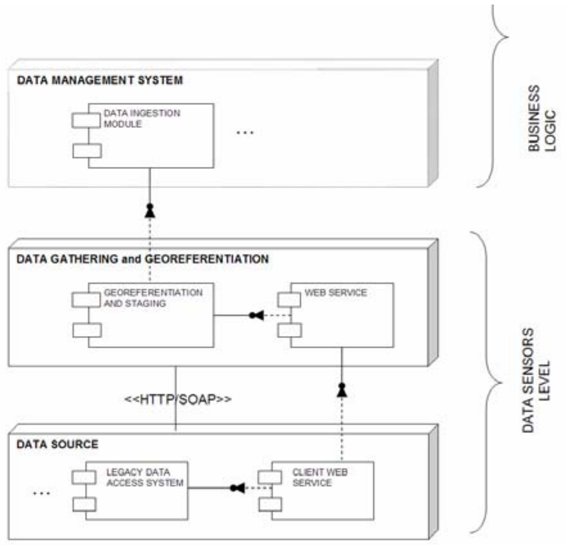

Figure 2. Detail of Data Sensors Level.

In detail publication interface will be made through web services [4] in order to divide client from server side implementation. In fact, web service is a set of web published functions that maybe used by other client applications. A web service provides standard methods for discovery and invocation services. Web services (as well as web applications) use HTTP protocol for communication and they are thought and modelled as web applications that produce HTTP responses for HTTP requests, other web services provide SOAP responses to HTTP/SOAP requests (SOAP - Simple Object Access Protocol – is a protocol built over XML language, it allows to invoke web functions and to get the responses). Web services specification information are provided through its WSDL (Web Service Description Language) that codify its invocation modalities and related response.

This level will include a module named *Data Gathering And Georeferentation* that collect and georeferencial the data with the chosen Spatial Reference System. Figure 2 shows in detail the interaction between a generic data source and this module.

3.2 Business Logic

The business logic level will include all modules that will manage from raw data staging to output models publishing. For this reason, it's necessary to develop three sub-modules:

a. **Data Management System**: it will be responsible for data reception related to: gathering modules, cataloguing and archiving modules. Particularly, cataloguing will be performed based on a common metadata set. The module will be responsible for managing climatologic archives too. The main module components will be the *Data Ingestion Module* (which will receive data coming from sensors, and will forward these to data catalog, and catalog data in accordance with available metadata), the *Data Catalog* (this will be the input of the relative Data Base Management System), *Data Search Module* (that will provide high level functionalities for data search inside archives);

b. **Resources Management System**: this is a sub-component that will provide the models execution and monitoring services on grid resources. It will provide all functionalities related to application management. Connection with users and grid resources will be realized through web services. In this level all core Grid services will be placed such as *Grid Information System* (responsible to interface with available computational grid resources to extract needful information for Jobs management), *Scheduling and Job Management* (responsible for job scheduling and submission) and finally the *Application Management* (responsible for high level requests management to execute models);

c. **Nowcasting Modules**: represent a container of modules, forecast weather and rainfall models. It will manage data processing, data fusion operations, etc.. Container structure is designed through a plug-in approach that will allow to insert and to delete various components ensuring the smallest impact on system structure.

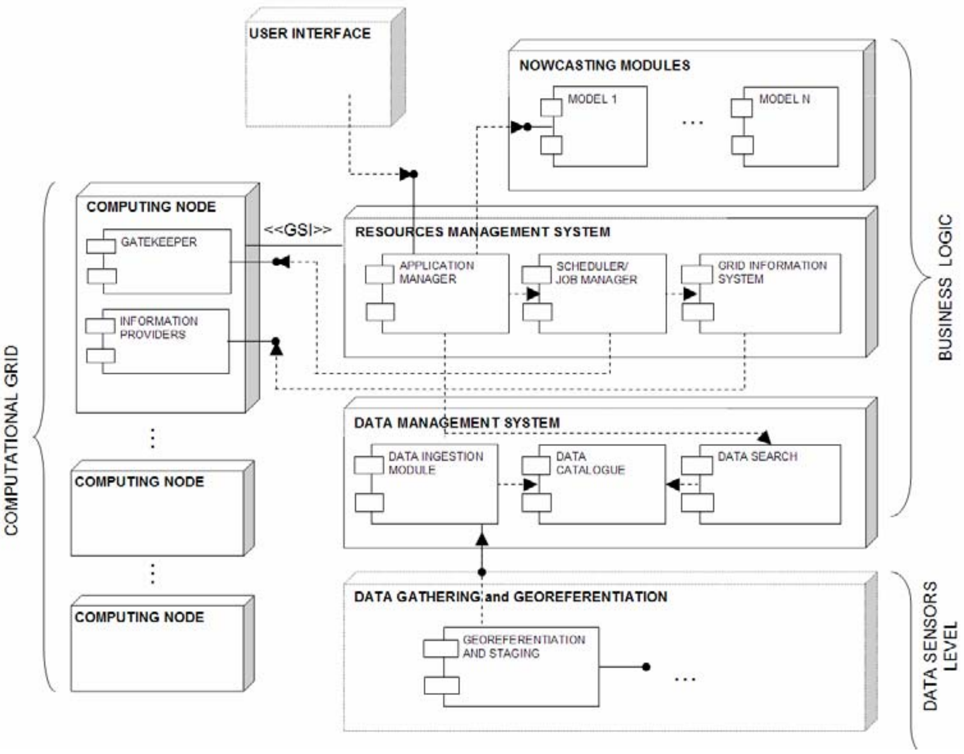

Figure 3 – Detail of Business Logic level

As schematized in Figure 3, Resources Management System interfaces itself to the computational grid through Scheduler/Job Manager (SJM) and Grid Information System (GIS) components. Computational grid is depicted by a computing node set. These components will ensure an adequate resources virtualization level because computing node numbers can vary dynamically, without alteration of proposed system architecture. GIS will make the available resources search and will provide status information about resources to the scheduler. Then, relatively to information provided by GIS, SJM will identify the resources on which the job will be submitted, according to adopted scheduling policy, it will start up file transfers, will submit the job and will retrieve the results. In fact, it will virtualize concern security mechanisms too, such as: on each grid's resource there are the necessary components to accept remote submissions and to provide information to GIS (such components are identified as Gatekeeper and Information Providers).

3.3 Presentation Level

It provides the interface set that will permit to obtain environment functionality access. Front-end acts as a data presentation level and than it will supply search services and output models display. It will be fully developed in accordance with web paradigm in order to minimize user client requirements: the users must only satisfy the requirements of a computer connected on Internet, with last generation web browser installation. This level includes the following components (see figure 4):

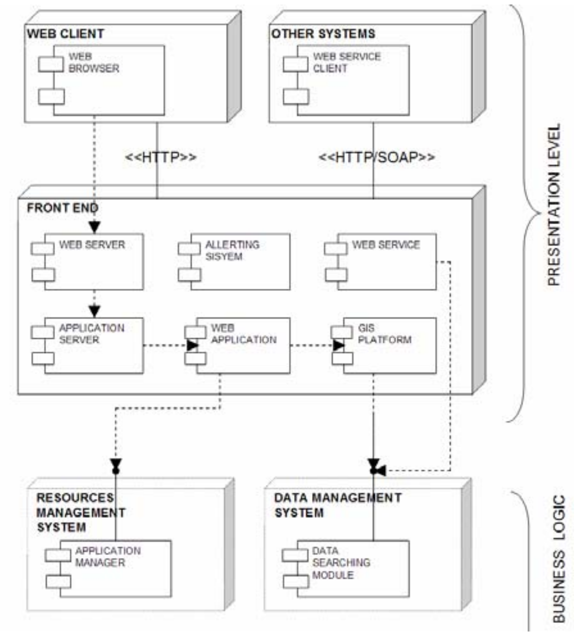

Figure 4 – Details of Presentation Level.

a. **Problem Solving Environment Interface (PSE):** this will be an interface component and it will provide the high level interactive functionality for problem definition and results visualization. In other words it will be a whole and integrated environment for simulation models composition, compilation and execution, based on available data. It will include advanced characteristics that aid users in problem formulation and execution. The component will include a configuration and knowledge container too. Container is developed through an ontology definition relative to: application domain, employed tools to work on container, and a semantic defined for tools description and possible data. This component allows the specialization of the environment to application context.

b. **Web services:** this is the machine to machine interface related to the output of models.

4 Discussion: Design Pattern and Technological Infrastructure

Proposed infrastructure will be developed through both Web and Grid technologies, because there are certain set benefits: first of all, this is a distributed system, and so we need to introduce intercommunication mechanisms between sensors and central system and between access/management positions and central system. De facto, by definition, web applications provides the possibility to increase or to make uniform access channels to corporate applications, introduce an extreme updating and distribution easiness. So, server centralized management, publication and/or updating and universal distribution united to multi-platform access (access is independent from hardware and operating system used by user, with some limits) are an undoubted benefit.

Design patterns used are the Model View Controller (MVC [5]) because it will allow an easy to trace and to expand system thanks to modular approach.

About base technologies, the middleware that will be used for computational grid implementation will be the Globus Toolkit [6]. Globus Toolkit is the standard de facto for computational grid environment.

Innovation elements produced by designed architecture related to infrastructural technology and methodology are the following:

Synergy of Sensors: the role of sensors (to measure atmospheric parameters in an advanced nowcasting system) is a matter of primary importance. In fact, the quality of yielded products is due to processing type and the added value is strictly related to synergies between several observation sources (radar, satellite sensors, conventional weather report data, etc.). The proposed methodological approach is strongly characterized in this way and represents a strong technological innovation. Particularly, for the first time in Italy, there will be a network of meteorological radars that will be unique in its kind. Innovative precipitation and cloud covering characterization products (originating from geostationary and polar satellites) are associated to meteorological radar network. The system uses data coming from non-meteorological sensors too. This is important for orography characterization, vegetation coverage and soil moisture. Finally, conventional data coming from ground weather stations are taken into account and aggregated.

Grid Computing Approach: the use of Grid computing is different from the existing infrastructures because of its scalability, reliability, robustness and performance. In fact, Grid computing allows efficient and secure aggregation of heterogeneous resources (computational resources, sensors, data, etc.) owned by different organizations. This is done by assuring an adequate security level. Moreover, transparent access to geographically distributed computational resources allows obtaining adequate processing times of images reducing costs too. Typical mechanisms related to resource management assures an adequate robustness of the system, thanks to advanced job management. The jobs are able to modify their executions on the grounds of alterations of resource status changes. Some reasons that lead us to choose such a technology are the following: Involved processing is often computationally intensive. As a consequence, the use of powerful computational resources reduces processing time drastically. High performance computational resources are limited therefore it is advantageous to share them between several organizations. Grid technology is the right choice with regards to secure sharing and coordinated use of computational resources; Satellite images (more generally remote sensed products) are often of large dimensions and are located in several repositories that are geographically distributed all over the world. It is calculated that several terabytes of data are acquired per month through earth observation satellites. Searching in these repositories causes several issues related to access, copyright and the confidential nature of the data itself. It is possible, through Grid computing, to realize systems, allowing transparent and efficient access to geographically distributed storages virtually; Satellite images are often processed following several steps in order to extract important information. Moreover, products coming from several sensors are integrated (data fusion) in order to extract more information about the same region of interest. Processing and integrating algorithms are often developed ad-hoc by specialists. Sharing of these algorithms is very advantageous for researchers and centers. Through Grid computing, it is possible to share software and knowledge too. This yields a collaborative environment enabling resource access transparently and use of tools developed by other users.

Specialized environment to the application domain: the approach used to realize the user front-end is typical of a Problem Solving Environment. This allows the proposed system to provide a set of access tools such as technological and implementative details which are hidden which to final users.

Possibility to publish outputs of models on several communication channels and to heterogeneous users: presentation level foresees a web service interface which assures the access to the output of models from other systems. Then it is possible to use the production in several fields, such as institutional sites, other research centres, etc. Finally, an automatic alerting system permits us to create prevention mechanisms at minimum costs. These mechanisms can be utilized, for example, by civil defence systems.

5 Conclusions

In this paper the design of a computational infrastructure for weather nowcasting, using Grid technology is presented. The paper aims to describe the design process with focus on the design choices and its implications on the final product. A discussion about the added value introduced by Grid technology is given. The described system is part of an approved Italian national research program that will start in the next few mounts. Future works will deal with the comparison between our system and the existing ones in the same applicative domain.

6 References

[1] Zawadzki I., “Nowcasting of precipitation”, World Weather Research Program Symposium on Nowcasting and Very Short Range Forecasting, Toulouse, France, 2005.

[2] I. Foster et al., “The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration”, Grid Forum white paper, 2003.

[3] Foster and C. Kesselman, “The Grid 2: Blueprint for a New Computing Infrastructure”, Morgan Kaufmann , 2003.

[4] Web Services Conceptual Architecture (WSCA 1.0), IBM, May 2001.

[5] Model-View-Controller [*http://www.ootips.org/ mvc-pattern.html*].