Grid Computing in the Context of Semantic Web-based Resource Management

Sabin Buraga, **Lenuţa Alboaie**
Faculty of Computer Science, "A.I. Cuza" University of Iaşi, Romania
{busaco, adria}@infoiasi.ro

Abstract
The paper draws several definitions and possible classifications of the Grid platforms, and enumerates the challenges and different areas of research. We expose the use of semantic Web technologies in the context of Grid computing.

Introduction

According to (Buyya, 2002), Grid computing – a new paradigm for next-generation computing – enables the sharing, selection, and aggregation of world-wide distributed heterogeneous resources for solving large-scale problems in certain areas of interest or for proving access to large data, information, or knowledge repositories.
Resource management and scheduling in existing environments is a complex undertaking. The geographic distribution of resources owned by diverse organization with different usage policies, cost models, and varying load and availability patterns is problematic. The producers (the owners of resources) and consumers (the users of resources) have different goals, objectives, strategies, and requirements.
To address these challenges, a systematic approach to model and retrieve certain resources has to be adopted. A system managing available knowledge must offer facilities for the creation, exchange, storage, and retrieval of knowledge in an exchangeable, platform-neutral and usable format (Singh, Iyer & Salam, 2005).
In this study, we describe the methods and technologies of semantic Web that can be used in the context of Grid computing for resource modeling/accessing.

Grid Computing: a General Presentation

Definitions and Characteristics

The actual Internet technologies' opportunities have led to the undreamt possibility of using distributed computers as a single, unified computer resource, leading to what is known as *Grid computing* (Buyya, 2002; Foster & Kesselman, 1999). Grids enable the sharing, selection, and aggregation of a wide variety of heterogeneous resources, such as supercomputers, storage systems, data sources, specialized devices (e.g., wireless terminals) and others, that are geographically distributed and owned by diverse organizations for solving large-scale computational and data intensive problems in science, engineering and commerce.

According to (Abbas, 2004; Laszewski & Wagstrom, 2004), different definitions of what Grid computing represent are available:

1. The flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources.
2. The capability to aggregate large amounts of computing resources which are geographically dispersed to tackle large problems and workloads as if all the servers and resources are located in a single site.
3. A hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to computational resources.
4. A way of processing distributed information available on Web.

One of the most used definitions is provided by (Globus): “Grid Computing enables virtual organizations to share geographically distributed resources as they pursue common goals, assuming the absence of central location, central control, omniscience, and an existing trust relationship”.

Virtual organizations can span from small corporate departments that are in the same physical location to large groups of people from different organizations that are spread out across the globe.

A *resource* is a shared entity available in the Grid. It can be computational, such as a personal digital assistant, laptop, workstation, server, and supercomputer or a storage resource. Other types of resources are the I/O ones: networks (e.g., bandwidth), printers, etc. Within a Grid, logical resources are also available: users, time counters and others.

Absence of a central location and central control implies that Grid resources do not involve a particular central location for their management. The final key point is that in a Grid environment the resources do not have prior information about each other nor do they have pre-defined security relationships (Abbas, 2004).

Architectural Model of the Grid

Grid applications are distinguished from traditional Internet applications – mostly based on client/server model – by their simultaneous use of large number of (hardware and software) resources. That implies dynamic resource requirements, multiple administrative domains, complex and reliable communication structures, stringent performance requirements, etc. (Buyya, 2002).

Some of the important issues regarding resource sharing across boundaries of organizations are (Abbas, 2004): *Identity and Authentication*, *Authorization and Policy*, *Resource Discover*, *Resource Characterization, Resource Allocation, Resource Management, Accounting/Billing/Service Level Agreement (SLA),* and *Security.*

To address these problems, a set of protocols and mechanisms need to be defined and/or adopted. The Grid architecture model is a layered one (Figure 1). Certain protocols, services, and APIs (Application Programming Interfaces) occur at each level of the Grid architecture model.

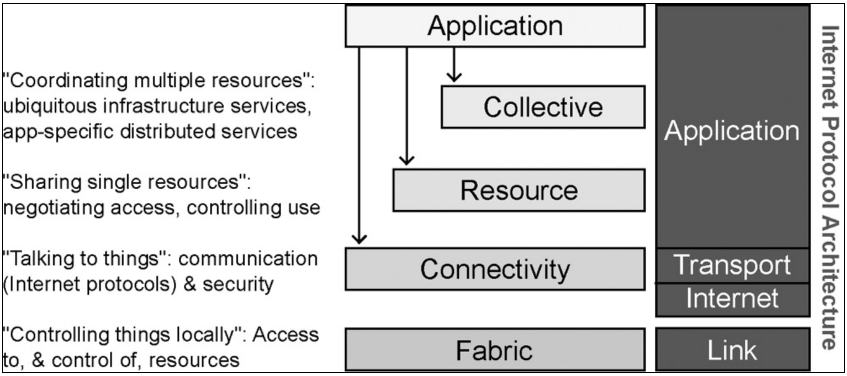

Figure 1. Grid model architecture (Abbas, 2004).

Grid Initiatives

Globus is a reference implementation of the Grid architecture and Grid protocols providing software tools in order to build grids and Grid-based applications – see (Joseph & Fellenstein, 2003). These open source tools are collectively called *Globus Toolkit* (current version is Globus 4). Other related projects are *Apple XGrid*, *Condor*, *Legion*, and *Sun Grid Engine*.

Open Grid Services Architecture (OGSA) concerns the use of Web services in the context of Grid computing (Berman, Fox & Hey, 2003). Grid services are in fact Web services (Erl, 2005; Alboaie & Buraga, 2006) executed to give access to resources by using actual Web technologies. To include different Web services extensions and to define stateful Web services, an important proposal is the *Web Services Resource Framework* (WSRF). The effort of standardization of Grid protocols, architectural models, and software tools is carried by the *Global Grid Forum* (GGF).

Several Classifications of Grid

Regarding the type of resources, a three layer model (Jeffery, 2000) for the Grid infrastructure was adopted by various research communities:

1. the *Computational Grid* – the lower layer concerned with large-scale pooling of computational and data resources that requires significant shared infrastructure to enable the monitoring and control resources in the resulting ensemble. This kind of Grid can be further classified in desktop grids, server grids, and high-performance/cluster grids (Abbas, 2004). A computa-

tional Grid can offer access to data or to compute resources (services), in order to augment other resources.

2. the *Information Grid* – this allows uniform access – via metadata (data about data) descriptions – to heterogeneous information sources and providing commonly used services running on distributed computational resources; the computational resources can vary, from simple method invocations to complete sophisticated applications;
3. the *Knowledge Grid* – provides specialized (meta-)services used for data discovery in existing data repositories and for managing information services; the meta-services aggregates many other types of services.

From the point of view of the structure of the organization (virtual or not) that uses a Grid platform, the following classification could be made (Abbas, 2004):

1. *Departmental Grid* – is deployed to solve problems for a particular group of people within an enterprise. The existing resources are not shared by other groups within the enterprise. This kind of grids is named *cluster grids* (in the vision of Sun corporation, represents one or more systems working together to provide a single point of access to users) or *infra grids* (term used by IBM to define a Grid that optimizes resources within an enterprise and does not involve any other internal partner; physical localization could be a campus or interconnected campuses);
2. *Enterprise Grid* – consists of resources spread across an enterprise and provides certain services to all users within that enterprise (also called *intra grid* or *campus grid*);
3. *Extraprise Grid* – is established between companies, their partners, and their customers, at the level of the extranet. These Grid resources are generally made available through a virtual private network, in order to reach a common goal;
4. *Global Grid* – is established by several organizations to facilitate their business or purchased in part, or in whole, from service providers. It is based on the public Internet technologies (especially, via a Web portal). This type of Grid is also called *inter grid* by the IBM.

Semantic Web and Grid Services = Semantic Grid Services

To achieve significant knowledge acquisition and management, a Grid system must support user collaborative tools and, more important, must provide a means of adding metadata (data about data) about the concepts and relations established between the resources within a given Grid platform.

When advancing towards Semantic Web (Shadbolt, Hall & Berners-Lee, 2006; Buraga, 2006), the main obstacle is the effort that the creator of information must put into organizing the knowledge and metadata, into tagging entities and relations, using vocabularies he must be familiar with, in order to make it comprehensible not only for humans, but also for machines.

The solution for this problem is to offer software instruments which help *transparently* organizing data and metadata for machine-comprehensibility. The users of the Grid applications do not need to know semantic Web vocabularies, as the system automatically generates metadata based on the creator's actions and on the progress of the information manipulated within the platform, in order to classify and to further retrieve given resources. From our point of view, a Grid system can be enhanced to support semantic Web technologies, in order to create, manage and present knowledge for any categories of users.

Additionally, Grid services can be semantically enriched by metadata and ontological descriptions. We can mention *Semantic Web Services Framework* (SWSF) which includes several proposals like *OWL-S* and *Semantic Web Services Ontology*. Another initiative is *Web Service Modeling Ontology* (WSMO).

Using semantic Web-based descriptions for Grid services, the applications will automatically discover, invoke and compose the desired services. With the help of the ontologies, the inter-operability and execution monitoring are also possible.

A Grid resource accessed via a Web portal interface can embed the following metadata (for example, different DCMI constructs about an image created during a blood-flow simulation on the Grid and supplementary annotated by the specialists). Using such of annotations, any Grid resource can be easily grouped for further processing.

A possible interesting aspects in resource modeling and structuring is denoted by the *Simple Knowledge Organization System* – SKOS (Miles & Brickley, 2005), which provides a RDF-based model for expressing the basic structure and content of concept schemes (thesauri, classification schemes, taxonomies, terminologies, glossaries and other types of controlled vocabulary).

An example of knowledge management Grid system, which exploits the role of metadata and ontologies, is our proposed *tuBiG* system (Alboaie & Buraga, 2003; Buraga & Alboaie, 2004).

Conclusions and Further Work

In this paper, we presented the most important aspects about Grid computing. We briefly displayed the actual semantic Web technologies and connected them to the Grid interests.

Our focus is to identify/propose a conceptual model for specifying Grid resources by using Semantic Web actual standards and technologies, following (Alboaie & Buraga, 2003). The first step will investigate the possibility to attach a metadata level to a given Grid infrastructure. These metadata specifications will be the basis for more complex higher levels, such as an ontology level, which will address different problems in modeling of semantic Grid applications and their interactions. At a superior level, diverse assisting collaborative, coordination, and searching instruments have to be integrated into the Grid, in order to easily give access to knowledge.

References

A. Abbas (Editor), *Grid Computing*, Charles River Media, 2004.

L. Alboaie, S. Buraga, *Web services* (in Romanian), Polirom, 2006.

L. Alboaie, S. Buraga, S. Alboaie, "tuBiG – A Layered Infrastructure to Provide Support for Grid Functionalities", International Symposium on Parallel and Distributed Computing, IEEE Computer Society Press, 2003.

F. Berman, G. Fox, T. Hey (Editors), *Grid Computing*, Wiley, 2003.

S. Buraga, *XML Technologies* (in Romanian), Polirom, 2006.

S. Buraga, L. Alboaie, "A Metadata Level for the tuBiG Grid-aware Infrastructure", International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Mirton, 2004.

R. Buyya, "Economic-based Distributed Resource Management and Scheduling for Grid Computing", *PhD Thesis*, Monash University, Melbourne, Australia, 2002.

T. Erl, *Service-Oriented Architecture: Concepts, Technology, and Design*, Prentice Hall PTR, 2005.

I. Foster, C. Kesselman (Editors), *The Grid: Blueprint for a Future Computing Infrastructure*, Morgan Kauffmann Publishers, 1999.

K. Jeffery, "Knowledge, Information, and Data", *A Briefing to the Office of Science and Technology*, UK, 2000: http://www.itd.clrc.ac.uk/ActivityPublications/239.

J. Joseph, C. Fellenstein, *Grid Computing*, Prentice Hall PTR, 2003.

N. Shadbolt, W. Hall, T. Berners-Lee, "The Semantic Web Revisited", *IEEE Intelligent Systems* 21(3), 2006.

R. Singh, L. Iyer, A. Salam, "Semantic E-business", *International Journal on Semantic Web and Information Systems*, 1(1), Idea Group, 2005.

G. Von Laszewski, P. Wagstrom, "Gestalt of the Grid", *Tools and Environments for Parallel and Distributed Computing*, John Weley & Sons, 2004.

* * *, *Global Grid Forum*: http://www.globalgridforum.org/.

* * *, *Globus*: http://www.globus.org/.

* * *, *Semantic Web Services Framework*: http://www.daml.org/services/swsf/1.0/overview/.