V

Virtual Enterprise Environments for Scientific Experiments

Andrea Bosin
Università degli Studi di Cagliari, Italy

Nicoletta Dessì
Università degli Studi di Cagliari, Italy

Maria Grazia Fugini
Politecnico di Milano, Italy

Diego Liberati
Consiglio Nazionale delle Ricerche, Italy
Istituto di Elettronica e Ingegneria dell'Informazione e delle Telecomunicazioni Politecnico di Milano, Italy

Barbara Pes
Università degli Studi di Cagliari, Italy

INTRODUCTION

An electronic marketplace (*e-marketplace*) is a promising architectural model to develop collaborative supply chain management and integration platforms (Premkumar, 2003). The e-marketplace philosophy is based on a collection of economically motivated companies that need to cooperate with each other by exchanging data, resources, and competences. To this aim, competitive technical and economical solutions have been found by integrating resources from diverse organizations into a *virtual enterprise* (VE) (Travica, 2005), enabling groups of companies to jointly develop and commercialise products and services, which an individual participating company could not realise by itself due to limited technical and financial capabilities.

Just to mention a few advantages, a VE benefits from formalisation of existing cooperation, from integration of large enterprise processes, from access to complementary competencies, and from technology re-deployment. In a VE, employees, suppliers, and customers need to share common business processes and applications and need to work with a huge volume of information. Unlike single user applications, VE environments must not only meet the needs of many separate individuals but also of entire groups, each one is interested in pursuing specific business goals.

Analogously to what business managers do, scientists today should routinely rely on computers and information sharing over the Internet. In fact, nowadays, scientific experiments are supported by activities, which create, use, communicate and distribute information, whose organizational dynamics are similar to processes performed by distributed cooperative enterprise units. Indeed, scientific progress necessitates international collaboration; scientific scenarios such as the human genome project and particle physics experiments would not be feasible without automated collaboration. This scenario is in some sense similar to that of enterprise environments, whose progress requires large-scale cooperative processes and efficient access to very large data collections and computing resources (Pollock, 2004).

On this premise, the aim of this paper is to illustrate methods and tools for executing cooperative scientific experiments through a platform allowing for the integrated use of heterogeneous applications and software tools that were not designed specifically to promote interaction and cooperation, but still are inherently suitable for cooperation support. Specifically, the paper describes a framework for an *Advanced Lab for Bioinformatics Agencies* (ALBA) supporting the design and management of cooperative scientific experiments. By assuming bioinformatics laboratories as virtual enterprises, the framework defines the responsibility of enterprise units in offering scientific services and the set of rules under which the virtual enterprise performs a cooperative experiment. By discussing some specific classes of bioinformatics experiments, the paper describes their mapping into a service-oriented architecture for distributed scientific environments.

BACKGROUND

Currently, entire vendor organisations, ranging from the development team to the marketing and sales forces, can benefit from methods, components, and solutions based on information and communication technology (ICT) networked environments; recent research work investigates the required extent and the means to apply the VE paradigm successfully in order to reduce transactions and development costs, to shorten time to market, to limit shared risks, and to provide rapid adaptation to market needs.

On the other hand, recent improvements in network technology accelerate the adoption of the VE paradigm not just in the business context but for an entire class of collaborative scientific experiments that may come from a range of areas including the physical and life sciences, engineering, mathematics, environmental science and medical imaging/diagnosis.

The concept of "scientific experiment" is rapidly changing in an ICT oriented environment, moving from local laboratories towards networked applications. Today scientific experiments are mainly "data driven," since they use heterogeneous software systems to discover knowledge from a variety of information and data sources derived by the interaction with physical devices. These kinds of scientific experiments are continuously generating lots of data that are often very distributed: a networked computer is not a simple technical support, but rather an integrated part of an experiment.

The common practice is to employ a range of accessible methodologies ranging from toolkits specifically aimed at supporting experiments to generic software tools (MatLab, Mathematica, SciRun, NetSolve, Ninf, Nimrod, etc…) and more widely deployed infrastructures such as the Web and the *Grid* (Berman, 2003; Foster, 2003). It is important to observe that the Web is about exchanging information while the Grid is about sharing resources such as data and information repositories, experimental facilities and application services. Because of their need for high-performance computing resources, many scientists are drawn to grid computing as the infrastructure to support data management and analysis across organizations (De Roure, 2004a). The use of advanced computing technologies to support scientists is usually denoted as *e-science* (De Roure, 2004b).

The consequence is a radical change in design for experimentations where the collaboration is essential for combining approaches, combining skills, and sharing resources. On this basis, a key success factor to promote research intensive products is the vision of a large scale scientific exploration carried out in a networked environment, with a high performance computing infrastructure, for example, of cooperative *e-services* (Pilioura, 2001; Mecella, 2002) (of Web or Grid type), that supports flexible collaborations and computation on a global scale (Bosin, 2006b). The availability of such a virtual cooperative environment should lower barriers among researchers taking advantage of individual innovation and allow the development of collaborative scientific experiments.

The definition of a virtual e-science environment poses many hard technical challenges deriving from the need of data access and integration, and the scale, heterogeneity, distribution and dynamic variation of information. As shown in Table 1, VE environments and e-science environments have many elements in common, but the question is if and how existing VE models can support distributed scientific experiments.

On this premise, we aim to provide a framework (ALBA) involving the use of networked resources and tools to support distributed scientific experiments; each of one is modeled as a business process (*e-process*) formalized by a *workflow*.

Table 1. Elements common to virtual enterprises and e-science environments

Resource sharing
- *Data, Information, Knowledge* - *Computers* - *Scientific Instruments* - *Networks*
Large scale collaboration
- *Multi Institutional* - *Overlying traditional organisational structures* - *Large or small, static or dynamic*