# GRID-ENABLED DISTRIBUTED COMPUTATIONS IN THE CHEMICAL PROCESS INDUSTRIES

Bill Karakostas, Centre for HCI Design
School of Informatics, City University, UK

and
Antonis Kokossis
Process and Information Systems Engineering
School of Engineering, University of Surrey, UK

*Abstract*

This paper presents a computing architecture suitable for the information information management and the demanding calculations of the supply chain. A grid architecture is proposed for distributed computations for simulation studies, integrated software applications, process synthesis, and large-scale optimisation. The proposed network features numerous advantages over traditional, monolithic solutions in terms of scalability, software re-use, security and distributed resource discovery and utilisation. The proposition is a novel example of how advanced distributed techniques and paradigms can be applied in the area of chemical engineering to support the distributed simulation of chemical processes.

## Introduction

Companies integrate previously diverse systems to make business processes faster and more efficient. Boundaries are dissolving between enterprise resource management suites (ERP), supply chain management software, and customer relationship management (CRM). For many users, the priority is on integrating what they already have rather than buying new systems. For chemical process industries, in particular, the integration of product and process related information imposes additional challenges and a need for transparent access to models and data. The paper presents an architectural framework for the systematic integration of systems related resources. The architecture is proposed in the form of Grids (Foster and Kesselman, 1998). These were invented in the mid-90's to denote a proposed distributed computing infrastructure for advanced science and engineering. They support the co-ordinated resource sharing and problem solving in dynamic and multi-institutional virtual organisations. Grids are envisaged as large collections of software and hardware resources interconnected in peer to peer networks used by mobile and stationary programs to discover, access and process various forms of information and data concurrently. They closely relate to software multi-agent systems that support distributed paradigms for efficient computing over the Internet and essentially can act autonomously and perform information processing related tasks on behalf of its user.

## Key Concepts and Technologies

The Grid is an infrastructure for enabling the integrated, and collaborative use of high-end computers, networks, databases, and scientific instruments owned and managed by multiple organisations. The Grid has been characterised by a 3 layer model which consists of (i) a computation / data grid, (ii) an information grid, and (iii) a knowledge grid. Whilst considerable advances have been made on computation/data grid, using technologies such as Globus (Foster and Kesselman, 1997) and Legion (Natrajan *et al,* 2001). Chemical engineering companies do not yet have a taste for such technologies. Even at a research level, there are no paradigms or testbeds of applications to explain and disseminate the better use of the technology. More than benchmarking and evaluation, however, the opportunities appear more clearly in the systematic development of knowledge and the creative management of the information assets of the industry. The paper contributes with a proposition of a framework that integrates the three layers of grid using agents and ontologies.

### *Ontologies*

Ontologies explicitly specify and conceptualise knowledge (an abstract, simplified view of the world). The ontology could also stand as a set of concept definitions that categorise things that exist or may exist in some domain. The term ontology is borrowed from the Greek philosopher Aristoteles who refers to it as a systematic account of what can exist or 'be' in the world. In the fields of Artificial Intelligence and Knowledge Representation the term refers to the construction of knowledge models which specify a set of concepts, their attributes, and the relationships between them.

The key ingredients that make up ontologies are a vocabulary of basic terms and a precise specification of what those terms mean. The ontology provides a set of well-founded constructs that can be leveraged to build meaningful higher level knowledge. An informal ontology may be specified by a catalogue of types that are either undefined or defined only by statements in a natural language. A formal ontology is specified by a collection of names for concept and relation types organized in a partial ordering by the type-subtype relation. When the knowledge of a domain is represented formally, the set of objects that can be represented is called the universe of discourse.

Our approach is based on the systemtic development of ontologies to capture the concepts and vocabularies of the supply chain alongside the common language used by chemical engineers to design processes. This represents a major advantage for low experience users of the grid portal who will then be able to share in the wider knowledge of the community. The paper supporst ontology based derivatives of the XML language such as RDF and XOL to capture the domain semantics (Decker and Melnik, 2000).

### *Agents*

Search agents search the Web for relevant knowledge sources according to some ontology. Agents incorporate processing strategies for finding and querying knowledge sources, react to discovered knowledge, and make decisions based on the knowledge they have gained (Knoblock and Ambite, 1997). Agents are therefore comprehensively knowledgeable in a given domain, know how to communicate with other agents, know how to glean information from HTML-like documents on the Internet, and are able to locate sources of structured knowledge. Search agents can also support the automated evolution of ontologies.

## Architecture

The grid architecture (see Figure 1) presented in this paper brings together technologies in process systems engineering (simulation, optimisation, data-mining), GRID computation, agents, and ontologies in a demonstration of solutions to a challenging industrial problem. Specifically, it supports:

(i) Prototyping a grid-enabled infrastructure designed to function as local, company, and/or discipline-wide test-bed. The prototype embraces geographically distributed CPUs, grid-enabled versions of commercial engineering software, databases, in-house models and a set of services to enforce security, Quality of Service (QoS) and empower control as to the information shared and distributed.

(ii) How to probe the potential of the grid to sustain applications and integrate heterogeneous sources of information. A Web portal is proposed as an interface with the Grid, to register applications, and make ready for the disposal of intelligent agents to automate computations and promote discovery of information from within the grid.

(iii) Enabling features for knowledge and information discovery and mining.

## Grid Components

(i) Classes of computation and the data grid

Engineering computations are most complex and are expected to become even more demanding with a challenge to integrate different models, bridge scales or apply analysis along the entire supply chain. Challenges are more evident in the case of:

(a) *large-scale optimisation*: supply chain formulations invariably lead to large-scale optimisation problems that are inflated by the number of products, destinations, intermediate storage and distribution/delivery points. Conventional approaches will fall short to apply over-the-self techniques without sophistication in the distribution of calculations and without enabling the systematic use of data and information to accelerate and converge. The additional need to handle flexible and reliable operations require repetitive solutions over uncertain regimes and the parallel analysis of what-if scenarios and case studies.

(b) *Model and data integration*: applications whereby the challenge is much less on the size of the models but more on the massive information to analyse, pre-screen and integrate with the models to solve the problems. Strategic planning, scheduling and planning applications are typical examples with substantial potential for financial and logistic gains. Such applications are also most distributed and diversified as there is an obvious lack of structure in the information at hand or enough consistency with the information required or delivered by the models.

(c) *holistic approaches* based on the vertical integration of models: multi-scale approaches involve the integration of process synthesis with process simulation models, the integration of lump-parameter models with CFDs, or the integration of empirical models with models derived from first principles. The co-ordinated experimentation of ideas would also require the combination of simulation and optimisation, whereby strategic and conceptual models are combined with superstructure-based approaches and various types of simulation models. Holistic approaches further relate to sustainable aspects of the supply chain and the application of life-cycle principles.

(ii) Information grid

The lack of openness between business units, sites and industrial communities is partly attributed to prevent information from being used as a lever against sharing partners. The grid development requires prior registration of applications, models, databases, operation profiles, and designated servers that support commercial software and middleware. Access control lists can be safely designed to protect information leaks and safeguard sensitive know-how. The information flow is secured through ACLs designed to suit the purposes of business units or sites. The development can be further customised to “trade” information on a pay-per-use basis. Middleware constitutes bespoke software to coordinate distribution, pass data from one CPU to another, prepare information for the models, and aggregate or disaggregate information for a particular purpose.

Users have rising expectations on software investments to maximize the profitability of interactions with customers and suppliers, by combining information from financial and supply chain systems. Workers in factories, warehouses and call centres need better information about costs, so that they can make more cost-effective and profitable decisions. The grid enables an easy and coordinated access to the registered components, namely application software, models and data. Within the grid environment, the servers hosting software applications - or the CPUs that undertake the computation - could remain in the background of calculations. The users can, or even guided to, trace the software without prior knowledge of the location of servers.

(iii) Knowledge grid:

Grid networks are not only venues for ample computing power but also natural grounds for an intelligent processing of information into knowledge. The knowledge paths would be accounted by ontologies to relate the structure of the underlying information. At a more basic level, such paths would be static and repeat integration patterns in the minds of users. Basic integration task could involve the:

(i) search for software or data over the grid: users can mine operational patterns, prepare data to set optimisation and simulation software, and incorporate data-driven intelligence into the solvers

(ii) automated integration of applications: several stages of execution and data usage could be combined in the form of a single computational task. Stages could involve the aggregation of data to apply strategic planning (using software A), followed by a review and reports of the results (using software B), and subsequent optimisation of several site schedules (using a third piece of software C).

More exciting scenarios emerge with the use and development of business intelligence applications. Such developments would be able to sense changes in information patterns, monitor trends in markets and demands, and better prepare the manufacturing process for modifications. Discovery of new concepts could hint a need to adjust production, suggest markets or the development of new products, suggest the establishment of new manufacturing sites, explain the need for capital investment, or suggest the need for the better management of an operation.

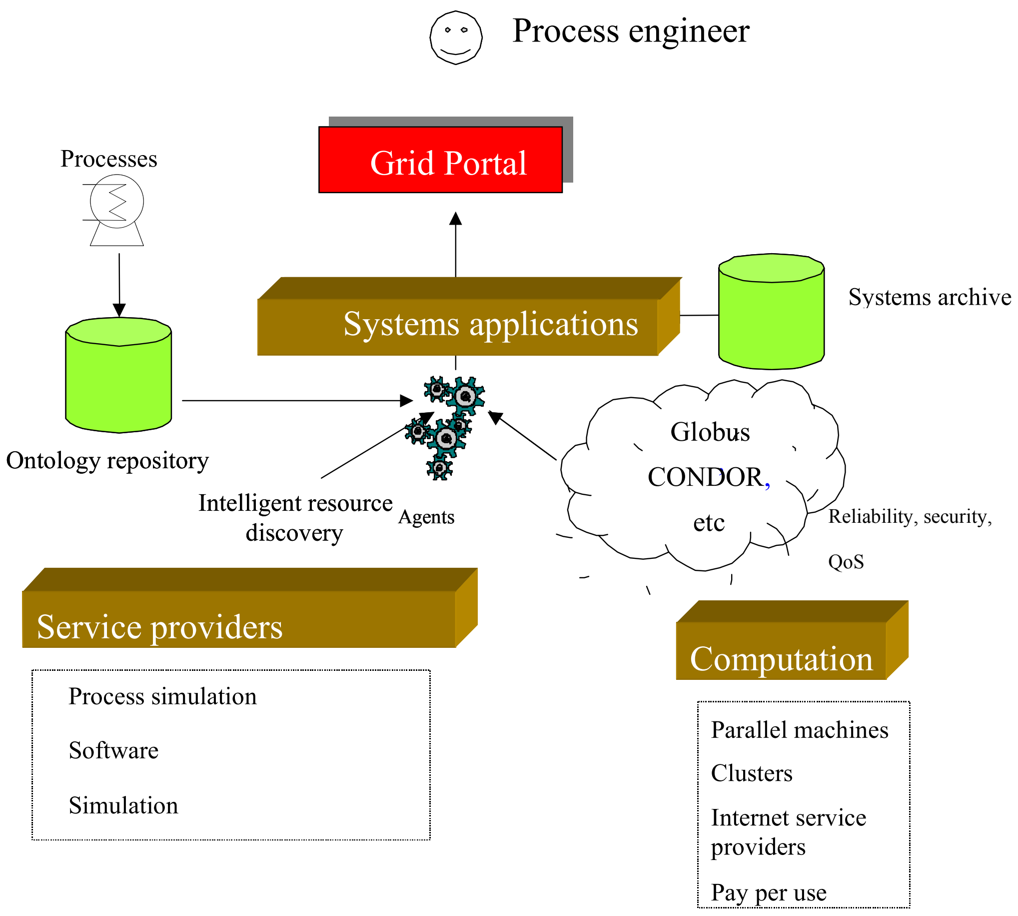

*Figure 1: Architecture outline*

## References

Decker S. and S. Melnik (2000) The Semantic Web: The Roles of XML and RDF. *IEEE Internet Computing*, **4**(5):63-74.

Foster I. and C. Kesselman (1998) *The Grid: Blueprint for a New Computing Infrastructure*. 550 pages ISBN 0-97028-467-5

Foster I. and C. Kesselman. (1997) Globus: A Metacomputing Infrastructure Toolkit. *Intl J. Supercomputer Applications*, **11**(2):115-128.

Knoblock C.A. and J.L. Ambite (1997) Agents for Information Gathering. In *Software Agents*, J. Bradshaw ed., AAAI/MIT Press, Menlo Park, CA.

Natrajan A, M. Humphrey, and A. Grimshaw (2001) Grids: Harnessing Geographically-Separated Resources in a Multi-Organisational Context. Presented at High Performance Computing Systems, June 2001