3rd International Workshop on Grid Computing for Complex Problems

GCCP 2007 Proceedings

Associate action to create national Grid initiative:
Making the Grid accessible for electronic science in Slovakia
(Sprístupnenie Gridu pre elektronickú vedu na Slovensku)

October 22 – 23, 2007
Bratislava, Slovakia

The workshop is organized by

Institute of Informatics, Slovak Academy of Sciences
Faculty of Electrical Engineering and Informatics, Technical University of Košice

The workshop is supported by

EU 6FP RI project - Enabling Grids for E-sciencE EGEE II (2006-2008) INFSO-RI-031688

Program Committee

Ladislav Hluchý	IISAS - Institute of Informatics, Slovak Academy of Sciences
Ján Paralič	Faculty of Electrical Engineering and Informatics, TU Košice
Ján Kollár	Faculty of Electrical Engineering and Informatics, TU Košice
Ján Sarnovský	Faculty of Electrical Engineering and Informatics, TU Košice
Milan Šujanský	Faculty of Electrical Engineering and Informatics, TU Košice
Karol Matiaško	Management Science & Informatics, University of Žilina
Ivan Hanuliak	Management Science & Informatics, University of Žilina
Jozef Černák	Faculty of Science, P. J. Šafárik University, Košice
Ľudovít Molnár	FIIT Slovak University of Technology in Bratislava
Milan Kolesár	FIIT Slovak University of Technology in Bratislava
Ladislav Hudec	FIIT Slovak University of Technology in Bratislava
Igor Hantuch	FEI Slovak University of Technology in Bratislava
Stojan Russev	Faculty of Economic Informatics, UE Bratislava
Claude Picard	CEREN - Centre d'Essais et de Recherche de l'Entente, France
Frederique Giroud	CEREN - Centre d'Essais et de Recherche de l'Entente, France
Veronique Cesari	Alter-Géo, La Lande, Cenac, France
Viet Tran	IISAS - Institute of Informatics, Slovak Academy of Sciences
Ján Astaloš	IISAS - Institute of Informatics, Slovak Academy of Sciences
Miroslav Dobrucký	IISAS - Institute of Informatics, Slovak Academy of Sciences

Organizing Committee

Ladislav Hluchý
Miroslav Dobrucký
Jolana Sebestyénová
Oľga Schusterová
Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia
E-mail: {hluchy.ui, dobrucky.ui, sebestyenova, sekr.ui}@savba.sk

Proceeding Editors

Ladislav Hluchý
Jolana Sebestyénová
Peter Kurdel
Miroslav Dobrucký
Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia
E-mail: {hluchy.ui, sebestyenova, peter.kurdel, dobrucky.ui, }@savba.sk

ISBN 978-80-969202-7-3

Preface

Welcome to the 3rd International Workshop on Grid Computing for Complex Problems GCCP 2007. The workshop is a two-day combined event for grid users: workshop with invited lectures, plenary discussions, accompanied by course for administrators of EGEE Grid sites, which is in the scope of EGEE II project – Enabling Grids for E-science 2006-2008, INFSO-RI-031688.

The topics of the workshop are:
- Grid and Service-oriented Computing
- Distributed Computing and Large Scale Applications
- Distributed Computing and Large Scale Data Access and Management
- Use of Knowledge and Semantics in Distributed Computing
- Grid Tutorials.

Special Session on Environmental Applications and Distributed Computing is devoted to the topics:
- Environmental Modelling, Simulation, and Optimization
- Environmental Risk Assessment, Planning, and Policy Making
- Remote Sensing and GIS.

The next goal of the workshop is an associate action to create national Grid initiative "Sprístupnenie Gridu pre elektronickú vedu na Slovensku" (Making the Grid accessible for electronic science in Slovakia) which will help to improve the e-Science in Slovakia through the creation of virtual organizations for individual science branches. The associate action aims to join Grid specialists with complex application users, to provide a medium for the exchange of ideas between theoreticians and practitioners to address the important issues in computational performance and computational intelligence towards Grid computing.

The workshop on Grid Computing for Complex Problems GCCP 2007 has attracted 25 paper contributions and active participations from France, Greece and Slovakia.

Many people have assisted in the success of this workshop. I would like to thank all the members of the Program and Organizing Committees, the workshop Secretariat for their work and assistance of the workshop. I would like to express my gratitude to all authors for contributing their research papers as well as for their participations in the workshop that made our cooperation more fruitful and successful.

Ladislav Hluchý
October 2007
Bratislava, Slovakia

Table of Contents

Invited lectures

Section 1
Grid and Service-oriented Computing

Section 2
Environmental Applications and Distributed Computing

Section 3
Distributed Computing and Large Scale Applications

Section 4
Use of Knowledge and Semantics in Distributed Computing

Section 5
Distributed Computing & Large Scale Data Access and Management

Invited lectures

High-throughput computing in molecular modeling and design

Tibor Kožár

Department of Biophysics, Institute of Experimental Physics,
Slovak Academy of Sciences, 04001 Košice, Slovak Republic
tibor@saske.sk

Abstract. High-throughput and high-performance computing has dramatically affected several basic and applied research areas. One of the areas that benefit from efficient computing is computer-aided molecular modeling. Efficient parallel and distributed computing has strongly impacted on molecular modeling in drug design and nanodesign related fields. This can be accomplished using versatile methodologies which can range from simple force field methods, high level *ab initio* quantum chemical calculations of molecular properties up to more complex large scale coarse-grain simulations of molecules and their assemblies. Parallelization of some programs started several years ago resulting in efficient high-performance tools. High-throughput computing on the recent grids facilitates calculations of interaction energies for millions of protein-ligand complexes. Virtual screening of large molecular libraries on the frame of the WISDOM projects [1] opened new horizons in computer-aided drug design combating neglected diseases. Herein, examples of methodologies employed in computer-aided drug design and nanodesign (and the overlap between them) will be presented.

Key words: High-throughput computing, molecular modeling, computer-aided molecular design, molecular dynamics simulations

1. Introduction

The impact of recent software development in molecular modeling, molecular simulation, chemo- and bioinformatics, along with hardware improvements and consequent cluster and grid technologies is remarkable. Computer-aided drug design (CADD) and/or computer-aided nanodesign (CAND) have both been significantly affected by high-throughput and high-performance computing. Nowadays, molecular modeling computations are the common core for both, CADD and CAND.

Several computation-based methodologies, such as molecular modeling [2], computational chemistry [3], virtual screening [4, 5], combinatorial library design [6], transition-state calculations, etc., have been confirmed to function as efficient and helpful tools during early stages of drug design. For example, the virtual screening methods are used for property-based evaluation of large compound libraries, filtering of adequate molecular properties and appropriate ADME/T (*A*dsorption, *D*istribution, *M*etabolism, *E*xcretion and *T*oxicology) profiles. Additionally, the virtual screening

The formation of the outer comet Oort cloud. Simulating the first giga-year of the evolution.

Marián Jakubík[1], Luboš Neslušan[1], Piotr A. Dybczyński[2], Giuseppe Leto[3], and Tomáš Paulech[4]

[1] Astronomical Institute, Slovak Academy of Sciences, 05960 Tatranská Lomnica, Slovakia,
mjakubik@ta3.sk; ne@ta3.sk
[2] Astronomical Observatory of the A. Mickiewicz University, Słoneczna 36, 60-286 Poznań, Poland,
dybol@amu.edu.pl
[3] Catania Astrophysical Observatory, Via Santa Sofia 78, I-95123 Catania, Italy,
gle@astrct.oact.inaf.it
[4] Astronomical Institute, Slovak Academy of Sciences, Dúbravská cesta 9, 84504 Bratislava, Slovakia,
astrotom@savba.sk

Abstract. Through the numerical integration of orbits of 10 038 test particles, which represent the initial proto-planetary disc, we follow their dynamical evolution during the period of 1 Gyr. We consider the perturbations by four giant planets, Galactic tide, and passing stars. The evolution results in the formation of the distant comet cloud known as the Oort cloud. We show that the population of the outer Oort cloud reaches its maximum at about 210 Myr. From about 500 Myr, it becomes almost stable, with only a moderate decrease. At 1 Gyr, the population decreases to about 40% of its maximum. The efficiency of the formation appears to be very low. Only about 0.29% of all considered particles reside in the outer Oort cloud at 1 Gyr. From about 50 Myr to the end of the simulation, the orbits are not distributed randomly, but high galactic inclinations of the orbital planes are strongly dominant. The dynamical evolution of the particles can be seen in several animations of their appearance in the space as well as in the animations of the evolution of several important quantities describing the process.

1 Introduction

In the process of the solar-system planet formation, small bodies were ejected, by the growing planets, into the interstellar space and large heliocentric distances. Those bodies, which have remained to be gravitationally bound to the Sun, formed a distant cloud. It is now known as the Oort cloud (OC) of comets. Since this reservoir has been little changed after the formation of the solar system, its dynamical study provides an unique opportunity to verify our knowledge of the cosmogony of early solar system.

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 17-24

Section 1
Grid and Service-oriented Computing

Web Services Composition Supported by Collaboration

Mária Bieliková, Ivan Kapustík, Pavol Návrat, Viera Rozinajová

Slovak University of Technology, Faculty of Informatics and Information Technologies
Institute of Informatics and Software Engineering
Ilkovičova 3, 842 16 Bratislava, Slovakia
{bielik, kapustik, navrat, rozinajova}@fiit.stuba.sk

Abstract. This paper describes our proposal to contribute to web services composition by support of collaboration. We divide this task into three levels – user level, model level and service level. On the user level we proposed a wiki-chat and semantic module, which support various ways of user communication and provide intelligent user support. Model level will be created as ontology description. Service level collaboration is designed as software agent collaboration. To improve the power of proposed agent system we decided to extend it to cover all three levels of collaboration.

Keywords: web services, service composition, collaboration, agent system, crisis management

1 Introduction

Web services are suitable for tasks where data sources and processing are highly distributed. One such area is crisis management and decision support for various fields such as for the radiological, environmental, hydrological or seismological emergence realized by the web information systems. Data are taken from the environment with typically larger scope than one country. Crisis management is realized by support of the models that simulate real situations. Required models are simulated on computers (often by means of grids due to their computational complexity on one side but rather sound decomposability) in specialized centers. The users (mainly the authorities responsible at local, regional or national levels for the crisis management) work often in groups – committees, where every member of the committee is responsible for specific part of crisis management.

It is obvious that emergency situations can vary in their type, location and other important features. This causes data sources, models, computer power and other resources to vary appropriately too. When we have all the processes realized as web services, the main task is to define right and efficient composition of the available web services. Design and development of the workflow composition of web services based on semantic description is active research area today. It is also one of the goals of the SEMCO-WS research project, which represents framework for research described in this paper.

Dynamic runtime environment and its usage for workflow submission*

Marek Kočan

Faculty of Science, P.J.Šafarik University,
Jesenná 5, 040 01 Košice, Slovakia
marek.kocan@upjs.sk

Abstract. For a job submitted to the grid, standard libraries and utilities that are distributed with the operating system are rarely sufficient. Further software is usually required. That is why the Runtime Environments have been introduced to grid computing. Usually these Runtime Environments need to be installed and maintained manually, but there are already solutions that are trying to automate the procedure of installation and maintenance of the Runtime Environments. These solutions can be then used also for submitting complex workflows from grid. This paper briefly presents one of the solutions for automate installation and maintenance of Runtime Environments and describes how Taverna as Runtime Environment can be used for submitting workflows from grid.

1 Introduction

The complex jobs submitted to grids rarely require just standard libraries and utilities that are distributed with the operating system. To solve this requirement the runtime environments [5] were introduced to grid computing. Grid runtime environments provide user interfaces to application software (and some other) resources in a way that is independent of the details of the local installation of the application and computing platform (OS, hardware, etc.). They are typically required by large research groups or user bases, dealing with a common set of software. The limitation of these runtime environments is, that they in most cases need to be installed and configured manually on each site and it is up to site administrators whether some runtime environment will or will not be installed on their site. That reduces the resources available for particular job.

On the other hand there is an application that allows users to integrate many different software tools, including web services, further it allows to design and execute the workflows, it is called Taverna workbench [7]. The limitation of Taverna is, that it allows executing just one workflow at the same time. In this paper we present a way how to pass this limitation by introducing dynamic runtime environments.

* This work is funded by the EU project KnowARC (FP6-032691) and project RPEU-0006-06 of Agency Supporting Research and Development (APVV).

Concentrator placement problem solution in the optimized radio network design

Jozef Petrek

FEI STU , Ilkovicova 3, 812 19 Bratislava, Slovakia
jozef.petrek@stuba.sk

Abstract. The goal of this paper is to introduce an algorithm for optimized repeater station placement using a digital terrain profile. Not only are the terrain profile and the radio link range included into consideration, but also the power source availability. The algorithm can be used to design radio-communication access networks. The basic optimization criterion is network cost and secondary also network reliability is taken into account. So a cost effective and computer supported design of radio network topology became possible.

1 Introduction

Computers are used at radio access network design for a long time [1]. A typical design tool uses a digital terrain profile where geographic coordinates are mapped in one of cartographics systems (e.g. WGS 84 or S 42) or artificial cartesian coordinates are introduced. Radio network design tools make usually possible to edit coordinates of transmitter and receiver stations, change their antenna height and to calculate the signal coverage or link profile. The microwave link is feasible if the first Fresnel zone is free, or else the signal is vastly attenuated [2]. The tools are usually able to display the first Fresnel zone and the link profile, while the radio earth curving is incorporated into the calculation. The radio network topology design itself must be done by the men (network designer). The algorithm described in this paper should replace the network designer and optimize the makes possible to optimize the network topology using a network topology design tool. This approach may save a big amount of money in comparison with “hand made” solution [3]. In [3] and in many similar works the optimization was made only for networks where the direct visibility of stations is not required e.g. in networks using optical cables or rented lines. In such case only the link length and the link capacity are all parameters taken into account to calculate the link cost. The cost function, which is a function of link cost and the link length, is rather simple. When a radio link is used the situation is much more sophisticated. In case of a radio link the cost function may be written only for the case of direct visibility between both stations while being rather complicated. When the visibility is not guaranteed or the link length is higher than the maximal possible one, an existing node must be used as a relay station. If such node does not exist a new node must be added into the network. This will increase the network cost dramatically because of rent cost, mast cost and radio technology (antenna, electronics, back-up battery…)

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 38-44

A Practical Component Framework for Development of Scientific Grid Applications

V. Šipková, R. Forgáč, J. Astaloš, V. D. Tran, and M. Dobrucký

Ústav Informatiky, Slovenská Akadémia Vied
845 07 Bratislava, Dúbravská cesta 9

Abstract. Software Component Frameworks are widely used standards in commercial business applications. In the last decade this technology is being explored with great interest as a way to build large-scale scientific applications on parallel computers and Grid systems. Nowadays, professional programmers attempt to build complex applications by composing the elements from large collections of predefined and tested units which are made available from several teams of specialists or open source community. The component technology fits very well with the manifold service-oriented Grids, however, the model must allow for a very dynamic control of composition. This paper describes a new component-based framework providing the higher-level components built on top of modern Grid technologies, that allow the easy and efficient composing and deployment of *Parameter Sweep Applications* onto a Grid platform.

1 Introduction

The computational science community hold huge existing investments in a broad assortment of physics, chemistry, climate, materials, fusion, combustion, radio astronomy, numerical systems, and visualization software. A scientific breakthrough is feasible by combining the best-in-class technologies from different disciplines into an integrated application. This is in practice not trivial to achieve due to codes are using different programming languages, data models, or different standards.

Software component frameworks are widely used standards in commercial business applications. In the last decade this technology is being explored with great interest as a way to build large-scale scientific applications on parallel computers and Grid systems. Due to the high computational complexity involved professionals attempt to build complex applications by composing the elements from large collections of predefined and tested units which are made available from several teams of specialists or open source community. This effort began in 1995 when a small group of collaborators tried to define the concept of the component architecture for scientific applications running on massively parallel systems. Their activity has grown to the establishment of the consortium [1] dedicated to defining and promoting the component specification standard called *Common Component Architecture* (CCA) [2]. The CCA model is brought about as a means to allow applications to run as components and easily interoperate

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 45-53

Workflow composition service for Grid applications

Viet D. Tran, Ladislav Hluchy

Institute of Informatics, Slovak Academy of Sciences
Viet.ui@savba.sk

In this paper, we present an approach for automatic workflow composition of Gird applications according to input and output data of tasks in workflow. The automatic workflow composition will reduce the work of application developers and allows dynamic and flexible workflow management.

1. Introduction

Over the last years, the development and acceptance of Grid technologies have been forwarded incrementally. Grid technologies connect distributed computational resources of dynamic multi-institutional virtual organization together and provide aggregate computational powers for solving very complex and computation demanding problems. The technologies make the infrastructures for researchers to share resources and knowledge, allows them to collaborate on solving common problems.

Since the infrastructure is becoming more and more powerful each year, the Grid applications also grow in size and complexity. The computation of the applications usually does not consist of a single task but many tasks connected together by data dependences. Workflow management became one of the main focuses of research and developments in Grid computing.

Although many workflow management tools are available, they have a common weakness. The users must prepare the workflow in advance using some text or graphical editors, and then submit the workflow to Grid for executions. If the workflows are complex, the process of editing the workflows may be time-consuming and error-prone for users. Furthermore, the approach does not allow modifying workflows during their executions: the users have to cancel the running workflow, use editor to modify the workflow description and submit it again. For applications with conditional workflows, where the actual workflows depend on the output from previous tasks, this approach is very inefficient.

This paper is organized as follows: in Sections 2, some existing workflow managers are reviewed. Section 3 introduces the approach of automatic workflow composition and Section 4 shows the context of automatic workflow composition tool in connection with existing middleware and task execution tool.

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 54-58

Section 2
Environmental Applications and Distributed Computing

Improvement of Fire Hazard Mapping : Road Traffic and Topography effects

Véronique Cesari [1], Frédérique Giroud and Claude Picard [2]

1 : Alter-Géo, La Lande, 24250 Cenac, France,
2 : CEREN, Domaine de Valabre, 13120 Gardanne, France.
e-mail: veronique.cesari@wanadoo.fr

Abstract— The processes involved in fire ignition are complex. Human and physical factors of fire ignition are here considered, to first determine the most relevant factors and then to combine them. In this study, fuzzy concepts which allow the integration of knowledge and the opinion of experts are used, to gather the large knowledge accumulated on fires in order to describe ignition mechanisms. As most ignitions occur in the vicinity of road networks, an assessment process for road traffic was developed in this frame using socio-economic variables to define population flows.

Index Terms— Fire hazard, fuzzy concepts, road traffic.

Introduction

Fire processes are complex because physical, meteorological and human factors are involved. Fire risk, or in our case fire hazard, is generally computed as an index (indexes reviewed in [1]), mainly based on meteorological data because they are the most critical variables for fire ignition and spread when an event occurs and to assess the background level of stress. Other hazard factors, physical and human factors, ([2],[3],[4] and [5]) are less considered in fire ignitions, even if they are important, as well as in ignition process characterization.
Fuzzy tools ([6]) allow the description of the complex fire ignition process and offer a realistic fire hazard assessment. The elaboration of factor maps and their combinations are defined to respect multi-source knowledge.
This study is based on previous results ([7]), improving factor analysis and takes into account road traffic through socio-economic estimations of population flows.

A visualization in cluster environment

Marek STRAKA, Branislav SOBOTA, Ján PERHÁČ

Department of Computers and Informatics
Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9, 040 01 Košice, The Slovak Republic,
Branislav.Sobota@tuke.sk
Marek.Straka@tuke.sk
Jan.Perhac@tuke.sk

Abstract. The paper describes a part of the virtual reality system PROLAND, that was developed on the Department of Computers and Informatics FEI TU Košice. This part is visualizing part, which is oriented to handling and visualization of large data sets on multi screen output devices. In first part is described a rendering in the cluster environment. The next part describes basic solution of hierarchy rendering model of visualization kernel which use multi-scene multi-screen mode. This model works in PC cluster and first practical results are described in the conclusion.

1 Introduction

Multi-screen display environments have been driven primarily by powerful graphics supercomputers, such as SGI's Onyx systems. Also the performance of graphics hardware is increasing at such a fast pace that a large class of applications can no longer utilize the full computational potential of the graphics processor. This is largely due to the slow serial interface between the host and the graphics subsystem. Recently, clusters of workstations have emerged as a viable option to alleviate this bottleneck. However, cluster rendering systems have largely been focused on providing specific algorithms, rather than a general mechanism for enabling interactive graphics on clusters. With features including shared-memory multi-processing and multiple synchronized graphics pipelines, they provided a stable and flexible development platform for high-performance virtual reality and visual simulation applications. Unfortunately, these features come at high cost. Hence, the use of multi-screen projection environments has been limited to a small number of users

During the past several years, high-performance and feature-rich PC graphics interfaces have become available at low cost. This development enables us to build clusters of high-performance graphics PCs at reasonable cost. An important issue, however, is that the programming model for shared-memory systems and clusters differ significantly. In shared-memory graphics systems, the programmer does not have to worry about issues such as sharing data amongst different processors or distributing rendering information to different graphics engines. In cluster environments,

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 68-73

Use of remote sensing data and GIS technologies for environmental modeling: The AIRSAT project

Panagiotis Symeonidis[1], Athina Kokonozi[1], Kostas Kourtidis[1], Evagelos Kosmidis[1]

[1]Demokritus University of Thrace, Dept. of Environmental Engineering
School of Engineering 67100 Kimeria-Xanthi, Greece
{symeonidis@draxis.gr, athinako@ceti.gr, kourtidi@env.duth.gr, kosmidis@dotsoft.gr}

Abstract. The present paper provides information about the rationale and implementation of the AIRSAT application. AIRSAT is an integrated, web-based satellite environmental information system. It provides to users environmental spatial data about meteorological parameters (temperature, humidity, pressure, wind speed, precipitation), atmospheric variables (O_3, NO_2, aerosols), and ocean geophysical variables (chlorophyll, sea surface temperature). Further, it provides online tools for covariance spatial data analysis for the geophysical variables included. AIRSAT thus provides for monitoring and studying the natural ecosystems, with focus on the atmospheric and oceanic environment, using remote sensing data, geographical information systems (GIS) and web databases and interfaces. The data that will be presented in AIRSAT are freely available for public access in internet and derive from various space agencies like ESA, DLR and NASA. AIRSAT, in its current configuration, is focusing on two geographical areas, one over Greece (local scale, with high resolution data) and one over Europe (continental scale, with lower resolution data).

Keywords: Remote sensing, GIS, Environmental Risk Assessment, ESA

1 Introduction

The last two decades, the atmospheric environment of Greece is under study not only in operational level with the creation of monitoring stations in urban and industrial environments for pollutants whose levels are controlled by directives and law, but also with research measurements of non-regulated pollutants from several Universities [1,2,3,4]. According to research of the last five years, transboundary pollution plays a significant role in Greek background concentrations (O_3, SO_2, aerosol) [5,6,7].

Currently, no operational monitoring of the background concentration (i.e. outside of urban and semi-urban locations) of pollutants in country level exists. On the Northern gateway to Greece, emissions from Central Europe and industrial installations in Balkans enter through Thrace. The contribution to local air pollution

3D Visualization Tool for Visualization of Running Grid oriented Applications

Eva Pajorová

Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia
eva.pajorova@savba.sk

Abstract. The duration of the computations in the Grid environment being exceedingly long, the user naturally wants to see the intermediate results and requires means to modify the running computations if the intermediate results are not satisfactory. This paper describes one of useful tools that provides means for run-time monitoring of the running Grid applications and displaying their intermediate results in the client application. Concrete describes 3D modeling and visualization tool for GRID computing natural disasters applications.

Key words: Grid computing, scientific visualizations, visualization service, virtual reality, natural disasters, flood modeling, fire modeling.

Introduction

One of the mentioned GRID_TOOLs is a tool for modeling and 3D visualization GRID – oriented applications of natural disasters, which was developed . It should integrate visualization requests of any kind of applications oriented on computing of natural disasters. In case it is grid computing it has to be established also submit workflow, which controls execution of this visualization service The natural disasters like fires and floods become subject of science in research institutions more and more frequently. Topic of a lot of projects is how to prevent such disasters. Many applications from this area are using different kinds of simulation tools, which are producing output data for displaying the results of the computation.

There are several tools solving the visualization of natural disasters. However, these tools lack common criteria for visualizations, unification of data formats, and a common approach to visualisation. The integration of formats and a common solution for visualization of natural disasters, which has been solved in our visual service, provided the opportunity to create a visualization service for grid applications aimed to natural disasters. As the applications computing natural disasters work with big amounts of data, grid computing is used frequently. Using visualization service could be considered a big contribution. This has been proved by increasing interest of experts working on international projects. Searching for new facilities of preventing natural hazards, they are continually seeking to integrate the existing applications. One of them is the visualization service with it's submit workflow.

Framework for Visualisation and Monitoring of Grid Applications

Peter Slížik and Ladislav Hluchý

Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia
peter.slizik@savba.sk hluchy.ui@savba.sk

Abstract. Typical Grid applications are usually very time-consuming. As the computational time increases, users naturally want to see the partial results that the application has generated so far, and modify the computation accordingly, if the partial results are not satisfactory enough. Furthermore, when submitting tasks to a geographically distributed computational infrastructure, users want to know where their applications are actually running. This paper describes a unified framework that provides means for run-time monitoring of running Grid applications and displaying their partial results in the client applications. The second part of the paper presents an overview of technologies that might be, in authors' view, the most efficient means for implementation of the framework.

Keywords: Grid computing, scientific visualisations, natural disasters, VRML, X3D, Ajax3D.

Introduction

Grid applications, by their nature, are very time consuming. Unfortunately, the existing Grid infrastructure supports particularly batch applications; support for interactive applications is only in its experimental stage. However, the time-consuming nature of Grid applications makes the users demand support for run-time checking and modification of behaviour of running applications. Two things are encompassed in this requirement: first, the user wants to have an overall view of the application's performance, and second, he wants to see the partial (incomplete) results the application has produced so far. In a distributed environment, another requirement is obvious: the user wants to see where (at which computation node) is his application being executed. Unfortunately, the current situation is far from optimal: the user has to wait until the whole application finishes and only then he can run the visualization job, which will present him his data in the form of imagery. Having to wait until the application finishes only to see the computation went wrong can be indeed very irritating.

Majority of Grid applications can be extended to contain some subsystem that would, on-demand or in regular time intervals, provide some information about the status of the computation; the typical examples of such information are the well-known progress indicator, or a computation phase indicator in case of multi-phase ap-

Section 3
Distributed Computing and Large Scale Applications

Tools for effective execution of parameter studies in EGEE grid

Ján Astaloš, Miroslav Dobrucký

Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 84507 Bratislava, Slovakia
{astalos,dobrucky}.ui@savba.sk

Abstract. Parameter studies are one of the most computationally demanding types of grid applications. They consist of many instances of the same simulation started with different input parameters or datasets. Usually all subtasks have to be finished to get the results. As the Grid environment is dynamic and not fully reliable, some tasks can fail due to various reasons – middleware errors, misconfigured grid sites, etc.. Sometimes tasks can end up in job queues and wait long for free job slots. All failed and waiting tasks are increasing overall runtime thus having negative impact on user's productivity. This paper describes design and prototype of tools developed to decrease the complexity of execution of parameter studies and to minimise the effort needed for their management. Use of the tools is demonstrated on astrophysical application.

Keywords: grid computing, parameter studies, fault-tolerance

1 Introduction

The demand for computing capacity by computationally intensive simulations led in recent years to creation of multi-domain geographically dispersed computing infrastructures, Grids[1]. One of the largest Grid infrastructure was built in the scope of EU project EGEE (Enabling Grids for E-sciencE)[2]. The infrastructure consists of more than 30 thousand CPUs and 5 petabytes of storage available to scientists from various application domains. One of the most computationally demanding types of grid applications are parameter studies (also called parametric simulations). They consist of many instances of simulation started with different input parameters or datasets. Since the runs are independent of each-other they can run on arbitrary number of computing nodes and thus exploit the computing power of computing

Material tension stress-strain curve determination via inverse analysis using finite element method in computational Grids – Conception of the analysis

Ladislav Écsi[1] , Pavel Élesztős[1], Viera Šipková[2] , Miroslav Dobrucký[2], Ján Astaloš[2]

[1] Department of Technical Mechanics, Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava, Námestie slobody 17
812 31 Bratislava, Slovakia
ladislav.ecsi@stuba.sk

[2] Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9
845 07 Bratislava, Slovakia
viera.sipkova@savba.sk

Abstract. The material tension stress-strain curve represents one of the most important material characteristics for elastic-plastic finite element analysis (FEA). While the curve can be determined quite accurately in small strain small deformation regions using standard tension tests, in large strain large deformation regions due to the loss of the Cauchy stress uniaxiality the accurate stress-strain curve determination is almost impossible. In this paper the authors present an alternative way of the aforementioned material characteristics determination using inverse analysis. The idea is based on a repeated FEA execution employing varying material stress-strain curve until the best results are reached. The calculation will be realized as a distributed model running in the Grid computing environment. In this paper the authors present some recent results from the ongoing work.

1 Introduction

From the mathematical point of view contemporary construction design has almost entirely used phenomenological approach based numerical calculations which utilize continuum and plasticity theory, combined with highly efficient numerical methods, such as the finite element method (FEM). The material stress-strain curve represents an important entry in these calculations and its accuracy essentially affects the numerical results. The curve experimental determination is quite straightforward in small strain small deformation region, but there are difficulties with it in large strain large deformation region. There isn't much work devoted to the area and there are quite a few publications either [4], [6]. In this paper we present an idea of the curve determination using an inverse analysis.

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 109-116

MATHEMATICAL MODEL OF AIRCRAFT AND ITS VISUALIZATION USING MPI

Peter KVASNICA[1], Tomáš PÁLENÍK[2], Michal ČIŽMÁR[3]

[1] Alexander Dubček University in Trenčín, Center of information technologies,
Študentská č. 2, 911 50 Trenčín, Slovak Republic,
kvasnica@tnuni.sk
[2] Alexander Dubček University in Trenčín, Faculty of Mechatronics,
Department of Informatics,
Študentská č. 2, 911 50 Trenčín, Slovak Republic,
palenik@tnuni.sk
[3] Exprit spol. s.r.o,
Švermova 23, 911 01 Trenčín, Slovak Republic,
michal.cizmar@gmail.com

Abstract. The article is denoted method of computing mathematic model in information systems of a simulator. For this reason uses information from control theory of design the decentralized systems created by computers. The concept arises from computers that create a distributed computer system of a flight simulator. This information system is created by computers and the program applications of the mathematic models. This modeling is accomplished by programming tool MPI, which are able to create a distributed computer system for a flight simulator. Mathematic modeling of the simulator system is made by equation depending on the architecture computer system. The important part of this article describes the solving mathematic models by different tolls. The analytical method of simulation and the analytical access of implementation parallel computing is realized one processor architecture.

The parallel computing at the computers of simulator illustrates time benefit and defined precision. This information allows in one time computing the mathematic models by MPI at the decomponated computer systems of the simulator.

Keywords: MPI (message passing interface), mathematic model, parallel task, flying simulator.

Metadata in Problem Solving Environment for Stereology Based Modeling

Július Parulek[1,4], Marek Ciglan[2], Branislav Šimo[2], Miloš Šrámek[1,3], Ladislav Hluchý[2] and Ivan Zahradník[4]

[1] Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia
[2] Institute of Informatics, Slovak Academy of Sciences, Slovakia
[3] Austrian Academy of Sciences, Austria
[4] Institute of Molecular Physiology and Genetics, Slovak Academy of Sciences, Slovakia

Abstract. This paper discusses the problem solving environment for stereology-based modeling applications. Such application involves tools for model creation, stereology-based model verification and model visualization. The application domain has complex and demanding technological requirements, including computationally intensive processing, operating platform heterogeneity and support for scientific collaboration. We exploit grid technologies to benefit of its high performance computing capabilities. Further, we are aimed at utilization of the metadata catalog for retrieval of previously computed models, which provides users a great advantage of grid environment.

1 Introduction

Recent progress in biological sciences, especially in morphology of biological objects, and the growth of computational power in computer sciences ask for development of geometrical modeling tools capable of creating 3D models of biological structures, which would make easier to grasp and communicate very complex features of biological objects. In the global concept of our work, we are aimed at better understanding of ultra-structure of muscle cells by means of electron microscopic studies and modeling. The structure of muscle cells has very specific features. Their volume is packed with numerous intracellular organelles of very complex three-dimensional organization placed within the intracellular proteinaceous gel, the cytosol. In the area of muscle cells, the structural properties are estimated by modern stereological techniques [1]. The list of properties includes shape, orientation, curvature, length, etc. of structural components. Highly desirable is to know the volume fraction and the surface area per unit volume of individual components in an entire object; also denoted as volume and surface densities. Biologists studying electron microscopic images of muscle cells are skilled in analysis and description of muscle cells according to these densities.

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 126-133

Section 4
Use of Knowledge and Semantics in Distributed Computing

Semantic description of the data mining grid services

Martin SARNOVSKÝ,[1] Peter BEDNAR[2], František BABIČ, Jozef WAGNER

[1]*Katedra kybernetiky a umelej inteligencie, Technická univerzita Košice*
Letná 9, 040 01 Košice
{martin.sarnovsky, peter.bednar, frantisek.babic,
jozef.wagner}@tuke.sk

Abstract. The Grid is considered as an integrated infrastructure for coordinated resource sharing and problem solving in the distributed environment. An ontology is an explicit specification of a conceptualization where definitions associate concepts, taxonomies, and relationships with human-readable text and formal, machine-readable axioms. The use of ontologies to describe Grid services will simplify and structure the systematic building of Grid applications through the composition and reuse of software components and the development of knowledge-based services and tools. In this paper we present the design and development of an semantic framework describing the Data Mining domain whose main goal is to semantically describe the particular services in order to simplify the development of distributed knowledge discovery applications on the Grid. The main goal of the proposed model is to offer to a user a reference model for the different kind of data mining tasks, methodologies and software available to solve a given problem, helping in finding the most appropriate technological solution.

Keywords: grid computing, semantic technology, service-oriented architecture, ontology modeling

1 Introduction

In many scientific and business areas, massive data collections need to be analyzed. Moreover, in several cases data sets must be shared by large communities of users that pool their resources from different sites of a single organization or from a large number of institutions. Grid computing has been proposed as a novel computational model, distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and high-performance orientation. Today Grids can be used as effective infrastructures for distributed high-performance computing and data processing [1].

An ontology is an explicit specification of a conceptualization where definitions associate concepts, taxonomies, and relationships with human-readable text and formal, machine-readable axioms. Ontology are used for :

User interface enabling smart semantic composition of web services

Emil Gatial[1], Branislav Simo[1], Martin Seleng[1], and Ladislav Hluchy[1]

[1] Institute of Informatics, Slovak Academy of Science,
845 07 Bratislava, Slovak Republic
emil.gatial@savba.sk

Abstract. This paper provides overview of proposed user interface and underlying portal framework that enables easy semantic composition of Web and Grid services. The interface serves for accessing and modifying ontological instances describing various service parameters, reviewing service composition, monitoring service run as well as for collaboration of users on some common task. The portal framework is developed within the SEMCO-WS project.

Keywords: Portal, JSR-168 portlets, workflow, user interface.

1 Introduction

SEMCO-WS application portal should provide a convenient web based user interface for the tools and applications being developed in the context of the SEMCO-WS project [1] (simplified component structure is shown on the Fig 1). SEMCO-WS portal builds upon the portal design experiences gained during the K-Wf Grid project and continue the development of more convenient user interface. The set of tools being developed in the SEMCO-WS project originates in the K-Wf Grid project with extended functionality or with implemented new features. The Users with valid portal account can access various portal sections focused on SEMCO-WS's applications and associated knowledge, application monitoring and grid services. The portal is built on the Gridsphere portal framework [2]. The framework is an implementation of a portlet paradigm defined in JSR-168 Portlet API Specification – portlets are self-contained components generating just part of the portal page, independent from each other that are pluggable into the portal. The Gridsphere framework provides the environment for running the portlets, portlet state management, session management and various other tasks. The gist of the SEMCO-WS portal development lies in the development of the portlets providing user interfaces to various software components.

A Distributed Semantic Repository for Web Services Workflow Composition

Ondrej Habala, Marek Ciglán, Ladislav Hluchý

Institute of informatics, Slovak academy of sciences
Ondrej.Habala@savba.sk

Abstract. The project SEMCO-WS[1], which started earlier in 2007, continues the efforts of the Institute of Informatics of the Slovak Academy of Sciences (II SAS) to design and develop software methods and tools for efficient composition of workflows of web services, on a scale encountered in major international computational grid infrastructures. II SAS specializes mainly in intelligent support of this process using ontologies, semantics, and reasoning. In this project, it is responsible for further developing the technologies for semantic web service integration, and also for a distributed ontology base, which would be able to seamlessly interconnect the various information and knowledge sources in the distributed middleware and application environment. This paper describes the main design goals, specification decisions, and preliminary analysis of such ontology base.

1 Introduction

The area of distributed computing has developed significantly in recent years. With the onset of grid computing before the turn of the millennium, a new era has begun, with the core of distributed processing system continuously shifting toward virtualization, inter-domain, and inter-organizational cooperation. Several large international projects have already tried to create a highly virtualized and homogenous infrastructure out of heterogeneous resources [1][3], and have been partially successful. The developed systems still more resemble a homogeneous and omnipresent grid of computing power, similar to the common electricity power grid [4].

The increasing size of projects, consortiums, and virtual organizations (VO), as well as the ambition to create reusable and interoperable software demands more expeditious standardization of communication protocols, module interfaces, information schemas, and user interfaces of computational grid middleware. For several years now, this standardization process is headed by the Global Grid Forum [5]. Initially, the standards for grid middleware, although using existing W3C [6] work, were developed separately from the business web services community [7]. Since this approach created an unwanted duplicity, the added features of the OGSA

[1] This work is supported by projects int.eu.grid EU 6FP RI-031857, SEMCO-WS APVV-0391-06, VEGA No. 2/6103/6, VEGA 2/7098/27.

Towards Semantic Annotation of Grid Services

Marian Babik, Emil Gatial, Michal Laclavik, Martin Seleng, Ladislav Hluchy *

Intelligent and Knowledge-oriented Technologies Group, Departement of Parallel and Distributed Computing, Institute of Informatics,
Slovak Academy of Sciences
Marian.Babik@saske.sk, Ladislav.Hluchy@savba.sk

Abstract. Web Service Resource Framework (WSRF) is a recent effort of the grid community to facilitate modeling of the stateful services [11]. Design and development of the WSRF service based systems is quite common and there are several emerging WS initiatives, which tries to automate the process of discovery, composition and invocation of such services. The semantic web services are a typical example, showing the potential of how ontological modeling can improve the shortcomings of the service oriented computing. One of the major obstacles in the process is the development of the ontologies, which describe web and grid services. Although, there are numerous standards for modeling semantic services, there are very few frameworks and tools, which can help automate the process of generating the semantic descriptions of services. This article presents a tool, which can semi-automatically generate the OWL-S descriptions for both stateful and stateless services based on the Web Service Description Language (WSDL) and corresponding annotations. Such functionality is inevitable in the grid environment hosting a vast number of services, which have to be semantically described in order to enable automated discovery, composition and invocation.

1 Introduction

Recently, Web service (WS) technologies are gaining importance in the implementation of distributed systems, especially grids. One such example is the Web Service Resource Framework (WSRF) [11], which extends the current WS technologies by modeling the stateful services. Design and development of the service oriented distributed system is quite common and there are several emerging WS initiatives, which tries to automate the process of discovery, composition and invocation of services. The semantic web services are a typical example, showing the potential of how ontological modeling can improve the shortcomings of service oriented computing.

In this paper we will introduce basic concepts of semantic web services (OWL-S) and web service resource framework (WSRF). Further, we will present the

* Acknowledgments: The research reported in this paper has been partially financed by Slovak national projects, NAZOU SPVV 1025/2004, SEMCO-WS APVV-0391-06, APVV RPEU-0024-06, APVV RPEU-0029-06, VEGA 2/6103/6, VEGA 2/7098/27.

Towards Pro-active Knowledge Sharing in Grid Environment*

Michal Laclavík, Martin Šeleng, Ondrej Habala, Branislav Šimo, Marek Ciglan

Institute of Informatics, Slovak Academy of Sciences, Dubravska cesta 9
845 07 Bratislava, Slovakia
laclavik.ui@savba.sk
http://ikt.ui.sav.sk

Abstract. Collaborative work in Grid environment is usually done by general purpose collaborative environments such as file sharing, groupware or wiki. We envisage the need for sharing information and knowledge in grid environment in a proactive way. By sharing information and knowledge users can collaborate and thus benefit. For such sharing of knowledge it is extremely important to deliver information and knowledge in context instead of searching for information by user. Thus we have created EMBET system in scope of successfully finished K-Wf Grid project. This tool is being extended to better detect context of user, context of knowledge and matching between these contexts. In this paper we discuss approaches for intelligent context matching, voting as well as other possible extensions of EMBET e.g. by Jabber chat server, thus discussing and describing ideas and work program to be carried in SEMCO-WS project regarding EMBET system improvements.

1 Introduction

Electronic tools for support of collaborative and cooperative work are gaining momentum quickly in last years, with the spread of broadband internet access and ingenious tools like the original WikiWikiWeb [12] or the recent Google family of web applications. Electronic – and especially internet-based – communication systems replace still more non-electronic services, beginning with the postal mail being largely replaced by the SMTP protocol, and ending with the establishment and success of the Wikipedia [13]. Many companies and projects develop and support frameworks [14][15][16] or complete solutions[17][18][19] which offer an integrated approach to communication and collaboration. Today, the information and collaboration technologies constitute one of the pilars of the so-called "intelligent enterprise" – the new model of successful enterprise management [20]. These technologies comprise of a large – and growing – set of electronic communication tools (e-mail, instant messaging, web publishing, or faxing), conferencing tools (computer-based forums, video conferencing, online chat rooms), and collaborative management tools (shared

* This work is supported by projects SEMCO-WS APVV-0391-06, RAPORT APVT-51-024604, VEGA 2/7098/27, NAZOU SPVV 1025/2004.

Section 5
Distributed Computing and Large Scale Data Access and Management

Fault tolerance techniques for grid environments: An overview.

Bartolomej Barra, Ladislav Hluchý

[1] Department of Informatics - Slovak Academy of Sciences
Dúbravská cesta 9, Bratislava

Abstract. Because of our present inability to produce error-free software, fault tolerance will continue to be an important consideration in software systems. The root cause of software design errors is the complexity of the systems. This paper presents an overview of fault tolerance techniques for grid environments. In the first part we show the basic overview of the software fault tolerance types (single-version, multi-version): Single version fault tolerance techniques include inline fault detection, checkpointing techniques and others. Multiversion techniques are based on the assumption that software built differently should fail differently, if one of the redundant versions fails, at least one of the others should provide an acceptable output: Recovery blocks, N-version programming, N self-checking programming, consensus recovery blocks, t/(n-1) techniques. An overview will be showned from existing research papers, how some of these techniques and algorithms are being used to achieve fault tolerance in distributed and grid environments.

Keywords: Fault tolerance, Grids, distributed computing, single-version, multi-version techniques

1 Introduction

For humans, perfect knowledge of a problem and its solution is rarely achieved. This means that software in general is hardly ever perfect. [2] states that “programs are really not much more than the programmer’s best guess about what a system should do”. The main problem is to build the essence of the software: the algorithms, data structures, functions and internal relationships between them.

Because of the impossibility to achive absolute certainty in software design correctness, the application of fault tolerance techniques in many cases is the only solution to achive the adequate stability, robustness and security in software systems; in our case in distributed computing systems and Grids. From the other point of view, there is a real possibility of hardware failures too, which grows by increasing the number of nodes and communication links between them.

Key Keeper service for grid data encryption

Marek Ciglan, Ondrej Habala, Ján Astaloš, Ladislav Hluchý

Institute of informatics, Slovak academy of sciences
Marek.Ciglan@savba.sk

Abstract. Data access security in compute grids is of high concern for multiple scientific communities. Enabling the data sharing in a controlled manner, under fine grained data access policy is a must. Today production grade grid middlewares provide highly secure methods for users authentication, however the access authorization methods are often insufficient, which is a major obstacle for certain communities to adopt the grid paradigm. In this paper, we propose a method for authorization virtualization from the concrete middleware and the enforcement of data owners' access policy over any storage system. The idea is based on the encryption of the data; the keys necessary to decrypt the data are available through specialized key-keeper service. The access policy enforcement can thus be performed at the level of the catalog service and is independent on the authorization methods used at grid nodes.

1 Introduction

Some of the data that are of interest for scientific communities are freely available, however in a lot of cases the data studied and processed by scientists and their applications must be protected by the highest security standards, as for example the medical data containing sensitive information about the patients.

The primary goal of the security systems is ensuring the authentication and the authorization. The authentication is the process of verifying the users identity. Once a user or software entity is identified the authorization mechanism is used to allow or deny the access to a grid resource.

Because of the distributed nature of grid systems, data security in grids is a non-trivial, rather complex problem. Unlike for the centralized system, the actors in grids must be managed and synchronized over numerous distributed resources. In addition, the constraint of grid resources autonomy must be kept. In other words, the site administrators and data owners must have absolute control over the authorization policies to their resources.

The authentication mechanism produce an user's identification token, let us call this token a distinguish name (DN). The distinguish name is used by the authorization mechanism to determine the privileges of the user. Authorization takes into account the DN and requested operation and verifies them against the security policies defined for a service or data items. In the scope data management, this means verifying the users privileges to access requested data item and to perform specified operation over the data item (read or write). Authorization is usually application specific procedure.

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 178-184

About Grid Workflows Verification

Peter Kurdel, Jolana Sebestyénová

Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
peter.kurdel@savba.sk, sebestyenova@savba.sk

Abstract. Grids are being adopted and developed in several scientific disciplines that have to deal with large-scale collaboration, massive distributed data, and distributed computing problems. National grid infrastructures in central Europe aim to join national and regional grid infrastructures and create robust and secure grid available to scientists. A Grid workflow system is a type of application-level Grid middleware that is supposed to support modelling, redesign and execution of large-scale sophisticated e-science and e-business processes. To ensure the correctness of grid workflow specification and execution, grid workflow verification and validation must be conducted so that we can identify any violations and consequently take proper action to remove them in time.

1 Introduction

Interactive computation [15] involves interaction, or communication, with the external world during the computation - in contrast to traditional, or algorithmic, computation, which proceeds in closed-box fashion. Concurrent and reactive computations, embedded, agent-oriented, distributed and component-based computations are also part of the interactive paradigm.

Service-Oriented Computing [24] utilises services as fundamental elements for developing distributed applications. One goal of Service-Oriented Architectures [10] is to integrate and compose services that are deployed on heterogeneous middleware paradigms. Web services encapsulate information, software or other resources, and make them available over the network via standard interfaces and protocols. Aggregating the functionality provided by simpler ones may create complex web services. This is referred to as service composition and the aggregated web service becomes a composite web service.

The service orchestration problem is a problem of making multiple services coordinate themselves and communicate in an orderly fashion so as to accomplish a task more complex than the single tasks provided by the individual composing services. [21]. The term workflow [2] can be defined as the orchestration of a set of activities to accomplish a larger and sophisticated goal. Significant research has been conducted in recent years to automate these activities using advanced workflow management tools. Some of the most popular and sophisticated workflow systems available in market include Websphere MQ Workflow [22], Staffware [20] etc. These products offer extensive functionality and support a variety of workflow patterns.

Proc. 3rd Int. Conf. Grid Computing for Complex Problems, Bratislava 2007, pp. 185-196

Job submission service based on WSRF[1]

Branislav Simo and Ladislav Hluchy

Institute of Informatics, SAS
{branislav.simo, hluchy.ui}@savba.sk

The paper presents the design and implementation of the Job submission service implemented on top of the Java WSRF standard implementation from Globus toolkit. The aim of this service is to provide preconfigured computational service to clients reaching it via network. Individual jobs are exposed as WSRF resources providing their state for reading and notification. Being easy to deploy on both Linux and Windows platforms the service is suitable for sequential short- to medium-lived computations that need to be deployed in a web service compatible manner.

Introduction

While there is a great demand for the powerful computational grids allowing executing arbitrary computationally intensive jobs, sometimes, however, there is a need for a fast application deployment in a web service manner. Such a deployment is useful when a scientist or developer of the application needs to allow other people to access it via internet without giving them physical access to the application or for easy web-enabling legacy command line applications.
In the MEDIGRID project [1], we were faced with the problem of creating architecture for deployment of legacy applications for easy remote execution. Applications were running either on Linux or Windows, some were parallel MPI applications, the others sequential running on workstations.
We had quite extensive experience with grid computing so the first candidate was to create a small grid infrastructure and run applications in it. That would require installation and maintenance of the grid software and porting the Windows based applications to Linux, as Windows was not supported as a target platform in the grid middlewares we had experience with – Globus Toolkit[2] and LCG[5]. The porting was not feasible and our experience showed the maintenance of full-fledged grid requires experienced administrators. Therefore we have decided to try to use web service approach and chose the Java WSRF implementation from the Globus to implement a job submission service that would allow us to wrap the legacy applications behind it and call the service transparently from both operating systems. That WSRF implementation already provided certificate based user authentication

[1] This work was supported by EU 6FP project MEDIGRID (FP6-2003-Global-2-004044), Slovak national project VEGA 2/6103/6, Slovak national project SEMCO-WS APVV-0391-06, APVV project RPEU-0029-06, EU INTAS Ref. Nr. 06-1000024-9154.

Author Index