1. K-WGrid - Knowledge Based Workflow System for Grid Applications

K-WfGrid Consortium

The K-WfGrid project which started on 1 September 2004, addresses development of a person-centric Grid environment, easy to program, to configure and to manage, based on standard software and protocols.

The Grid is a large, distributed collection of Grid or Web services which offer computing, data storage, networking, and specialized simulations. Due to the dynamic nature of the Grid, workflow composition is a challenging task because the system has to deal with resource unreliability and unpredictability, which is closely related to fault tolerance and scheduling.

In the framework of this Project we have developed and implemented (as a prototype) knowledge-based support for workflow construction and execution. The Grid system we have elaborated
- semi-automatically composes a workflow of Grid services,
- executes the composed workflow application,
- monitors the performance of the Grid infrastructure and applications,
- analyses the resulting monitoring information,
- captures the knowledge contained in the information,
- reuses the joined knowledge gathered from all participating users in a collaborative way in order to efficiently construct workflows for new Grid applications.

In this way, the results of K-WfGrid close the gap between monitoring and workflow construction.

K-WfGrid combines together and extends current achievements in Web Services, Semantic Web, Grid services and agent technologies to enable easier composition, execution and monitoring of workflows for real-life applications. The K-WfGrid system assists users in composing powerful Grid workflows by means of a rule-based expert system. The system adopts ontological descriptions of the environment and applications. We have developed a distributed knowledge base, a network of intelligent Grid services, as well as sophisticated performance monitoring and analysis, advanced user interfaces, and knowledge discovery.

The elaborated solutions were applied to three Grid applications: flood scenario simulations, city traffic pollution emission computations and enterprise resources management services.

The details are presented in around 110 research papers and on 17 K-WfGrid posters at CGW'06.

The K-WfGrid Consortium is coordinated by Fraunhofer Institute for Computer Architecture and Software Technology; Berlin, Germany, and, besides of this institution, there are 3 academic partners: Institute for Computer Science, University of Innsbruck, Austria, Institute of Informatics of the Slovak Academy of Sciences, Academic Computer Center CYFRONET of the AGH University of Science and Technology, Cracow, Poland, as well as two industrial partners: LogicDIS S.A. from Greece and Softeco Sismat SpA from Italy.

Acknowledgements
The described work is supported in part by the European Union through the IST-2002-511385 project K-WfGrid

References

[1] www.kwfgrid.eu

2. K-Wf Grid Portal - a Web Interface to Semantic Workflows

Emil Gatial(1), Branislav Simo(1), Giang Nguyen(1), Michal Laclavik(1), Ladislav Hluchy(1), Tilman Linden(2), Bartosz Kryza(3)
(1) Insititue of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
(2) Fraunhofer FIRST, Berlin, Germany
(3) ACC CYFRONET AGH, Krakow, Poland

The K-Wf Grid portal provides a central user interface, point of interaction and access to the components developed within the K-Wf Grid project [1].

The main functionality of K-Wf Grid portal is to provide users with access to the components developed within the K-Wf Grid project and other requested resources. This covers the components for workflow specification, workflow management and visualization, components dealing with the metadata management and components for accessing the other Grid services. Moreover K-Wf Grid portal includes collaborative feature provided by lightweight discussion portlet.

Architecture of K-Wf Grid portal is based on the Gridsphere portal framework [2] that provides an open-source portlet based Web portal. GridSphere enables developers to quickly develop and package third-party portlet web applications that can be run and administered within the GridSphere portlet container. The basics features of Gridsphere includes management of user accounts, groups and access to specific portlets, providing customized layout. Portlets provide "mini application" that can either display informational content or provide access to other components. Portlets can be assimilated into the portal web pages with customized layout.
K-Wf Grid portal includes many different components (portlets) that access the components of K-Wf Grid. One of the most relevant portlets is GWUI portlet that accesses GWES component [3]. It is implemented as a Java applet placed in the GWUI portlet. This portlet allows users to manage a semantic composed workflow by the well arranged workflow graph.
UAA portlet provides access to the UAA component [4]. Originally the interface is implemented as a servlet and it is wrapped by the UAA portlet. This portlet allows user to specify a problem and view the notes relevant to such problem. UAA component can advise the user on already composed workflows for similar problem. Component implements dynamic refreshment mechanism using AJAX technology.
GOM Browser portlet employs services implemented in GOM [5] and allows user to browse and manipulate the specified ontology stored in GOM. While the portlet manages access to GOM resources the visualization of ontology is mostly based on the Javascript (XML parser and AJAX) that moves the processing on the client browser. This speeds up the process of visualization and makes the notion of seamless application.
Data Browser portlet wraps servlet for accessing Grid services such as GridFTP transfer, RLS service. Using this portlet user can download or upload files from/to specified storage element. Moreover, the Data Browser can be used as a stand alone servlet accessed directly from URL request.
Logging portlet shows logging messages produced by the project components, thus making it possible to monitor inner workings of those components. It is mainly meant for component developers as a conveniet way of monitoring of their software and the cooperation of all the projects components.
Discussion portlet is lightweight collaborative portlet. User can create a new topics, answer to existing topics and view topic's threads or topics only. This porltet is integrated with UAA to annotate text of topics.

The main innovation of the portal is providing an uniform access to the components developed within the K-Wf Grid project with using the technologies such as AJAX for seamless communication between client and server. The interfaces are wrapped by the JSR-168 compliant portlet and deployed within the Gridsphere portal framework.

The paper describes detailed K-Wf Grid portal architecture along with the user interfaces of the key components developed within the K-Wf Grid project.

Acknowledgements. This research has been partially supported by the Project K-WfGrid (EU IST-2002-511385). The authors thank the K-WfGrid Consortium for many feedbacks during the design and development process.

[1] K-WfGrid Project - http://www.kwfgrid.eu
[2] Gridsphere web page - http://www.gridsphere.org
[3] Hoheisel, A.: User Tools and Languages for Graph-based Grid Workflows. In: Special Issue of Concurrency and Computation: Practice and Experience, Wiley, 2005 doi10.1002/cpe.1002 (in print)
[4] Laclavik M., Gatial E., Balogh Z., Habala O., Nguyen G., Hluchy L. Experience Management Based on Text Notes (EMBET) [PPT] In: Proc. of eChallenges 2005 Conference, 19 - 21 October 2005, Ljubljana, Slovenia, Innovation and the Knowledge Economy, Volume 2, Part 1: Issues, Applications, Case Studies; Edited by Paul Cunnigham and Miriam Cunnigham; IOS Press, pp.261-268. ISSN 1574-1230, ISBN 1-58603-563-0.
[5] B. Kryza, J. Pieczykolan, M. Babik, M. Majewska, R. Slota, L. Hluchy and J. Kitowski: Managing Semantic Metadata in K-Wf Grid with Grid Organizational Memory, In Bubak, M., Turala, M., Wiatr, K. (eds): Proceedings of the CGW'05, pp. 66-73, ACC CYFRONET AGH, Krakow, Poland, 2006

3. User Assistant: Towards Collaboration and Knowledge Sharing in Grid Workflow Applications*

Michal Laclavik, Martin Seleng, Zoltan Balogh, Ladislav Hluchy
Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia

In this paper we describe a User Assistant for collaboration and knowledge sharing in Grid Workflow applications. Theory, implementation and use of such system are described. The key idea is that a user enters notes in a particular situation/context, which can be detected by the computer. Such notes are later displayed to other or the same users in a similar situation/context. The context of user is detected from computerized tasks performed by user. Also intelligent components in the grid middleware such as monitoring, workflow analysis or workflow composition can provided context sensitive notes to be displayed for the user. User Assistant was created in scope of K-Wf Grid project. In the K-Wf Grid, grid services are semi-automatically composed to workflows, which should solve a user problem. It was identified that even when services and input and output data are well semantically described, there is often no possibility to compose an appropriate workflow e.g. because of missing specific input data or fulfillment of a user and application specific requirements. To help user in workflow construction, problem specification or knowledge reuse from past runs it is appropriate to display notes and suggestions entered by users or intelligent middleware components. Thus experts can collaborate and fill up application specific knowledge base with useful knowledge, which is shown to users in the right time.

References

1. HABALA O., BABIK M., HLUCHY L., LACLAVIK M., BALOGH Z.: Semantic Tools for Workflow Construction; In: Proc.of International Conference on Computational Science, Part III, LNCS 3993, Springer-Verlag, 2006, pp. 980-987, ISSN 0302-9743, ISBN 3-540-34383-0, May 28-31, Reading, UK.
2. Michal Laclavik, Martin Seleng, Marian Babik: OnTeA: Semi-automatic Ontology based Text Annotation Method; ITAT 2006, NAZOU Workshop, 26. 9 - 1. 10. 2006, Chata Kosodrevina, Bystrá dolina, Nízke Tatry, 2006
3. Michal Laclavik, Marian Babik, Zoltan Balogh, Emil Gatial, Ladislav Hluchy; Semantic Knowledge Model and Architecture for Agents in Discrete Environments ECAI 2006, The 17th European Conference on Artificial Intelligence, Riva del Garda, Italy, Aug 28th - Sept 1st 2006
4. Michal Laclavik, Martin Seleng, Emil Gatial, Zoltan Balogh, Ladislav Hluchy: Ontology based Text Annotation - OnTeA; 16th European - Japanese Conference on Information Modelling and Knowledge Bases (EJC 2006), Trojanovice, Czech Republic, May 29th - June 2nd, 2006
5. Michal Laclavik, Zoltan Balogh, Emil Gatial, Ladislav Hluchy; Collaboration and Knowl-edge Sharing in Grid Applications In: Proc. of 1-st workshop Grid Computing for Complex Problems - GCCP 2005, VEDA, 2006, pp.103-111, ISBN 80-969202-1-9. 29th November 2005 - 1st December 2005, Bratislava, Slovakia
6. LACLAVIK M., GATIAL E., BALOGH Z., HABALA O., NGUYEN G., HLUCHY L.: Experience Management Based on Text Notes (EMBET) In: Proc. of eChallenges 2005 Conference, 19 - 21 October 2005, Ljubljana, IOS Press, pp.261-268
7. LACLAVIK M., BALOGH Z., NGUYEN G.T., GATIAL E., HLUCHY L.: Methods for Presenting Ontological Knowledge to the Users In: L.Popelinsky, M.Kratky (Eds.): Znalosti 2005, pp.61-64. ISBN 80-248-0755-6. February 2005, Slovakia

----

* This work is supported by projects K-Wf Grid EU RTD IST FP6-511385, NAZOU SPVV 1025/2004, RAPORT APVT-51-024604, VEGA No. 2/6103/6.

4. Grid Workflow Execution Service - Dynamic and Interactive Execution and Visualization of Distributed Workflows

Andreas Hoheisel
Fraunhofer Institute for Computer Architecture and Software Technology, Berlin, Germany

The Grid Workflow Execution Service is the workflow enactment engine of the K-Wf Grid system [11], which coordinates the composition and execution process of Grid workflows. It implements a highly dynamic workflow concept [4][10] by means of the Grid Workflow Description Language (GWorkflowDL) [1][2][3], which is based upon the theory of high-level Petri nets [5]. It provides interfaces to the Web Portal and to a command line client for user interaction and to the low-level Grid Middleware for the invocation of application operations.

The main purpose of the Grid Workflow Execution Service and its client interfaces is to enable users to automatically execute workflows on distributed resources without being bothered with implementation-specific details, putting more attention to the functionality of the workflow itself. Therefore the Grid Workflow Execution service provides methods to initiate and analyze Grid workflows, and to coordinate and optimize the execution of these workflows on distributed and inhomogeneous resources regarding the control as well as the data flow. Abstract operations are automatically mapped onto matching software and hardware resources, triggering web service operations, remote executions of programs, or file transfers. The workflow service supports several kinds of Grid middleware, such as pure Web Services and Globus Toolkit 4, and it is easily extendible for further execution platforms, such as UNICORE, Condor, GRIA, and SUN Grid Engine, allowing reusing existing workflows on different Grid middleware.

The Grid Workflow Execution Service is deployed as a standard Web Service, so it seamlessly integrates into common system architectures. There is a strict separation of the enactment machine and its client interfaces. The graphical user interface is realized by a set of Java Applets and Servlets which can be easily integrated in generic or application-specific web portals. In addition there exists a command line client program that implements the full set of workflow management features and which simplifies the integration into legacy frameworks. The workflow synchronization between the execution service and the graphical user interface is done by means of an advanced protocol which makes use of the XUpdate syntax in order to transfer workflow modifications or status changes from the execution service to the clients and vice versa. The Grid Workflow Execution Service delegates parts of the refinement process from abstract to concrete (executable) workflows to external web services, such as the Workflow Composition Tool (WCT) the Automatic Application Builder (AAB), the Resource Matcher, and the Scheduler [6][7][9][13].

The main scientific achievements include the definition of a theoretically well founded Workflow Description Language, which is very simple but expressiveness at the same time, using the well-known concept of high-level Petri nets. In our concept transitions (symbolized by squares) represent workflow activities, such as web service operations or the remote execution of programs. Tokens (symbolized by filled circles) are used to represent instances of input or output data as well as side effects. Tokens are hold by places (symbolized by empty circles) which are connected with transitions by means of directed edges (arcs). With this approach we are able to model arbitrary algorithms, including parallel branches or loops. Transitions can possess conditions that we express using the XPath 1.0 syntax.
In order to make the workflows persistent, the execution service uses a native XML database for storing snapshots of the workflows including the workflow's state, which is represented by the marking of the corresponding Petri net. The system provides a simple and robust checkpoint-restore functionality and the user may initiate each of these snapshots as a new workflow instance.
Another scientific achievement is the overall concept of iteratively mapping abstract workflows onto concrete resources, taking into account the most current monitoring information. Each of the abstraction levels uses the same workflow description language, so it is possible to mix several levels of abstraction in one workflow.

The Grid Workflow Execution Service has successfully being applied in several application domains, such as city traffic management, environmental simulations [8] and media post processing. The research described here is supported in part by the European Union through the IST-2002-511385 project K-Wf Grid [11] and the IST-2002-004265 Network of Excellence CoreGRID [12].

References

[1] Martin Alt, Sergei Gorlatch, Andreas Hoheisel, Hans-Werner Pohl: A Grid Workflow Language Using High-Level Petri Nets. CoreGRID Technical Report Number TR-0032, Institute on Grid Information, Resource and Workflow Monitoring Services, CoreGRID, 2006
[2] Martin Alt, Andreas Hoheisel, Hans-Werner Pohl, Sergei Gorlatch: A Grid Workflow Language Using High-Level Petri Nets. R. Wyrzykowski et al. (Eds.): PPAM05, LNCS 3911, pp. 715-722, Springer-Verlag Berlin Heidelberg, 2006
[3] Martin Alt, Andreas Hoheisel, Hans-Werner Pohl, Sergei Gorlatch: Using High Level Petri-Nets for Describing and Analysing Hierarchical Grid Workflows. In: Proceedings of the CoreGRID Integration Workshop 2005, Pisa, 2005
[4] Tomasz Gubala, Andreas Hoheisel: Highly Dynamic Workflow Orchestration for Scientific Applications. In: Proceedings of the CoreGRID Integration Workshop 2006, Cracow, 2006
[5] Andreas Hoheisel and Martin Alt: Petri Nets. In: Workflows for eScience, Ian J. Taylor, Dennis Gannon, Ewa Deelman, and Matthew S. Shields (Eds.), Springer, 2006
[6] Andreas Hoheisel, Thilo Ernst, Uwe Der: A Framework for Loosely Coupled Applications on Grid Environments. In: Grid Computing: Software Environments and Tools. Cunha, Jose C.; Rana, Omer F. (Eds.), 2006 ISBN: 1-85233-998-5
[7] Andreas Hoheisel: User Tools and Languages for Graph-based Grid Workflows. In: Special Issue of Concurrency and Computation: Practice and Experience, Wiley, 2005
[8] Andreas Hoheisel: Workflow Management for Loosely Coupled Simulations. In: Pahl-Wostl, C., Schmidt, S., Rizzoli, A.E. and Jakeman, A.J. (eds), Complexity and Integrated Resources Management, Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society, Manno, Switzerland, Volume 1, pp. 500-505
[9] Andreas Hoheisel, Uwe Der: An XML-based Framework for Loosely Coupled Applications on Grid Environments. P.M.A. Sloot et al. (Eds.): ICCS 2003. LNCS 2657, 245-254, Springer-Verlag Berlin Heidelberg, 2003
[10] Andreas Hoheisel, Uwe Der: Dynamic Workflows for Grid Applications. In: Proceedings of the Cracow Grid Workshop '03. Krakau, Polen, 2003
[11] K-Wf Grid Project – http://www.kwfgrid.eu/
[12] Network of Excellence CoreGRID – http://www.coregrid.net/
[13] Falk Neubauer, Andreas Hoheisel, Joachim Geiler: Workflow-based Grid Applications. In: Peter Sloot (Ed.) Future Generation Computer Systems, Volume 22, Number 1-2, pp. 6-15, Elsevier, 2006

5. Workflow Composition Tool for Building Abstract Workflows

Tomasz Gubala(2), Daniel Harezlak(2), Maciej Malawski(1), Marian Bubak(1,2)
(1) Institute of Computer Science, AGH Univerity of Science and Technology, Krakow, Poland
(2) ACC CYFRONET AGH, Krakow, Poland

The Workflow Composition Tool is designed and developed in the K-WfGrid project [1] as a component of the workflow orchestration and execution environment. The aim of this tool is to construct a new abstract workflow as a solution to the input problem which is specified by a user as a set of results required from the future processing. The composer tool delivers a workflow of service operations that would eventually provide the specified results. The solution should be described with the terminology understandable for the user and it requires to be abstract enough to support subsequent reuse of solutions. Together with the rich set of tools that support refinement, scheduling and execution of workflows the composer forms an environment for modeling, planning and gathering results of complex scientific applications on the Grid [2].

The main workflow composition functionality is defined as a transformation of a user request document into a solution workflow document in the same notation format. The initial workflow description document that specifies all the results required by the user is provided in K-WfGrid's notation (GWorkflowDL) and it may contain some parts of workflow already specified and it may have another (yet unknown) parts to be constructed by the composer. The tool parses the initial workflow document in order to find all the parts that still need to be developed. Such part (called an unsatisfied dependency) usually contains specification of data that is required at a certain stage of workflow processing - thus the main objective is to find suitable providers for this data in form of one or more Grid services. The composer queries an external ontological registry and, as a result, an identification of suitable operations of Grid services is delivered. As the newly found operations may have their own dependencies (for instance in form of input data), the composer follows the same algorithm again to complete the workflow. When all the dependencies (both the original ones specified by the user and the ones introduced by new operations) are satisfied the workflow is complete. If there are no appropriate service operations available that would fulfill current requirements, the unresolved parts of the workflow are labeled and may be refined again when the environment changes. The final workflow is returned to the user in the same GWorkflowDL notation. In order to provide better request-to-service matching, the algorithms implemented in the tool, are based on semantic description of services' operations which are the building blocks of the composed workflow. With the matchmaking process based on Input-Output-Preconditions-Effect vectors, the composer is able to find the best available provider of the data requested by the user. The composer is accessible for the external users through a SOAP interface. Additionally, there is provided a simple set of servlet GUI for standalone use by a human [3].

The tool is designed as a service that offers its functionality through remote interface. This loosely coupled architecture enables easier distribution of components; one may consider the composer as a stateless transformation that stores all the information it requires in an external memory. As the local resources are not crucial for the tool, it may be freely relocated and replicated as long as it maintains a connection with the ontological store that contains registry of the available services. Internally, the tool contains three main functional blocks. The first one is used to retrieves every unsatisfied dependency from the workflow, the second one delivers a solution (e.g. with querying an external registry for suitable services or with searching the description of the workflow whether there are already appropriate services there). The third block incorporates the resulting solutions into the current state of the composed workflow.
Those three components are activated in subsequent iterations until the workflow is completed [2]. The tool uses widely accepted SOAP standard for communication with external systems and it relies on the recent achievements of semantic web to obtain and process semantic information (RDQL/ITQL, OWL). In order to increase the possible applicability the tool is able to cooperate with different services that provide ontologies (the Grid Organization Memory developed within the K-WfGrid Project and the open-source Kowari metastore). It also uses XPath query standard in order to retrieve reusable subworkflows document from the eXist XML database and to incorporate them as parts of constructed solutions.

The most important achievement of this development are the algorithms and heuristics that are able to combine theory of description logic of service operations with the strict formalism of PetriNet-based workflows. The graph-theory algorithms are needed to properly introduce new elements into a complex workflow graph and to add workflow patterns like XOR-join (the case when a result may be delivered through more then one operation) or AND-join (when an operation needs more then one input). The further methods search the composed workflow internally in order to check if a suitable operation is already a part of the workflow or not - this approach supports another useful processing patterns like AND-split (a provider of a given result is already used in the workflow for another purpose), XOR-split (more than one operations need exactly the same input to operate) and loops (a certain stage of the workflow has to be executed more than once). Special care is needed in such cases to provide these workflow patterns and still guarantee there are no deadlocks or undesired computation redundancy. In result, the implemented tool is able to construct complex and useful workflows which include more complex patterns like loops, parallel execution branches and deferred choice branches. Despite the complexity the composed abstract workflows are still valid. Another post-processing optimization algorithms assure that the amount of user-provided data is minimized and they also reduce the size of the resultant workflow graph to make it more legible to a human. As the tool uses domain-specific ontologies for workflow element description, the entire processing is defined in a vocabulary that is understandable to the end user (the ontologies provide common language in this case) [4].

This tool is an element of the K-WfGrid's workflow orchestration frameworkwhich enables a semantic workflow reuse. As the solutions proposed by the tool are abstract, their descriptions do not outdate fast, so they may be reused. The workflows may be stored in an XML document database and then semantically described in a dedicated workflow registry. The tool may be used to search through these reusable workflow with a similar technique as that applied for the operation discovery, and a found subworkflow is incorporated into currently composed one to form a new uniform description [5].

In the K-WfGrid Project, the Workflow Composition Tool was applied to flood scenario simulations, city traffic pollution emission computing, and to enterprise resources management [1].

Acknowledgements. This research was supported in part by the EU IST-2002-511385 project K-WfGrid, the IST-2002-004265 Network of Excellence CoreGRID, and the related SPUB-M grants.

References

[1] www.kwfgrid.eu
[2] T. Gubala and M. Bubak and M. Malawski and K. Rycerz: Abstract Workflow Composition in K-WfGrid Project Environment, in 4th Cracow Grid Workshop (CGW04) - Workshop Proceedings, pp.85-92, ACC Cyfronet AGH 2005
[3] T. Gubala and M. Bubak and M. Malawski: Support for Automatic Workflow Composition Based on Ontologies in Semantic Grid Environment, in 5th Cracow Grid Workshop (CGW05) - Workshop Proceedings, pp.34-40, ACC Cyfronet AGH 2006
[4] T. Gubala and M. Bubak and M. Malawski and K.Rycerz: Semantic-based Grid Workflow Composition, in 6-th Int. Conf. on Parallel Processing and Applied Mathematics PPAM'2005, pp.651-658, Lecture Notes in Computer Science vol 3911, Springer, 2006
[5] T. Gubala, D. Harezlak, M. Bubak and M. Malawski. Semantic Composition of Scientific Workflows Based on the Petri Nets Formalism. e-Science'2006 Conference, 4-6 December 2006, Amsterdam, accepted

6. Automatic Application Builder - Tool for Automatic Service Selection

Lukasz Dutka(1), Szymon Natanek(1, 2), Jacek Kitowski(1, 2)
(1) Academic Computer Centre CYFRONET-AGH, Cracow, Poland
(2) Institute of Computer Science AGH, Cracow, Poland

The global evolution of grid systems and trends towards breaking any barriers in the access to grid, makes the environment much more complex then it used to be. At the moment the grid systems evaluate into a gigantic set of grid or web services deployed across to world where every minute might bring new tools deployed as services. The trend makes obvious that the future grid applications will be workflows consisted of distributed services. On the other hand gird must be user-friendly and accessible for everybody. Thus in K-WF Grid[1] we developed infrastructure automatizing process of construction workflow-based application trying to provide application basing only on general description of expected results provided by the user of the application.

To make that possible we exploited semantic technology and knowledge processing at several stages of the application construction process. First of all basing on general description of the expected results provided by the users one of the K-WF Grid tool tries to find a way of providing expected by the user results without going into details - the result of this phase we call an abstract workflow. Then basing on current status of the system description gathered by a central K-WF Grid knowledge repository (called Grid Organization Memory - GOM[2]) Automatic Application Builder tries to automatically selects best services for the realization of the workflow task. And in the end the workflow is optimized from the performance point of view.

AAB uses extendable set of rules, which define criteria for searching the best fitted services in a specific context[3-4]. It is possible to use external source of information derived from external systems or services as a part of the rule set, such as GOM for ontological description of services, or any other tools provided vital information about the environment state. Moreover, AAB is a tool which strongly takes into account users' preferences or requirements which should be fulfilled by services being of the part of the final workflow-based application.

AAB is developed as web service. Such approach allows to improve accessibility to its functionality. It uses rule based expert system shell JESS [5] to realize reasoning engine.

The main achievements includes realization of rule based expert system with extendable set of rules which are processing information coming from many sources including users' preferences, external sources of information and so on. Described expert system was successfully integrated with K-Wf Grid.

The main scientific achievement is combining external sources of information inside of expert system, allowing to simplify execution of workflow operations and achieve better flexibility of the K-Wf Grid system.

The work described in this paper was supported in part by the European Union through the project K-WF Grid funded by the European Commission program IST-2000-25182.

References

[1] K-WfGrid Project - http://www.kwfgrid.eu
[2] B. Kryza, J. Pieczykolan, M. Babik, M. Majewska, R. Slota, L. Hluchy and J. Kitowski: Managing Semantic Metadata in K-Wf Grid with Grid Organizational Memory, In Bubak, M., Turala, M., Wiatr, K. (eds): Proceedings of the Cracow Grid Workshop '05, pp. 66-73, ACC CYFRONET AGH, Krakow, Poland, 2006

[3] L. Dutka, R. Slota, D. Nikolow, and J. Kitowski. Optimization of Data Access for Grid Environment. In 1st European Across Grids Conference, Lecture Notes in Computer Science, Universidad de Santiago de Compostela, Spain, February, 13-14 2003. Springer. (in print).
[4] L. Dutka, B. Kryza, K. Krawczyk, M. Majewska, R. Slota, L. Hluchy, J. Kitowski: Component-expert Architecture for Supporting Grid Workflow Construction Based on Knowledge. In Cunningham P. Cunningham M. (Eds): Innovation and the Knowledge Economy. Issues, Applications, Case Studies. pp. 239-246, Vol. 2, IOS Press 2005
[5] JESS - http://www.jessrules.com/

7. K-WfGrid Scheduler - Scheduling Grid Workflows Based on Prediction and Performance Knowledge

Marek Wieczorek, Thomas Fahringer
Distributed and Parallel Systems Group, Institute of Computer Science, University of Innsbruck, Austria

The K-WfGrid Scheduler is a performance oriented workflow Grid scheduler, responsible for making workflow applications ready for execution, by assigning workflow transitions to the resources where the web services corresponding to the transitions should be executed. It is meant as a tool which uses the performance estimation techniques (knowledge-based performance predictions and dynamic performance monitoring) developed in K-WfGrid project [1], and applies them to make a reliable workflow scheduling.

The Scheduler processes workflows created by the Workflow Composition Tool (WCT) and the Automatic Application Builder (AAB), which can contain multiple service deployments assigned to each workflow transition. It selects between semantically equivalent variants of web service deployments, assigning in that way transitions to the resources where the deployments are located. Multiple scheduling algorithms, implementing different graph traversing approaches (full-graph analysis or individual transitions analysis), supported by different types of auxiliary data (performance predictions or performance monitoring), can be used interchangeably.

In the K-WfGrid environment, the Scheduler acts as an auxiliary service for the Grid Workflow Execution Service (GWES), scheduling workflow applications on demand. The Scheduler uses two other components of K-WfGrid, called the Performance Monitoring and Analysis service and the Knowledge Assimilation Agent (KAA). These two services are responsible for dynamic performance monitoring and knowledge-based performance prediction (execution time prediction). Internally, the Scheduler is a single service which can use one of different algorithms implemented as interchangeable plug-ins. Currently, there are five algorithms implemented: two of them (easy and random) provide only basic functionality without any performance considerations, and three other advanced ones (performance-oriented, prediction-oriented, and prediction-and-performance-oriented) apply different performance estimation techniques. Two algorithms (prediction-oriented and prediction-and-performance-oriented) apply full-graph workflow analysis, whilst the other ones schedule each workflow transition separately (individual transition analysis). The easy algorithm chooses always the first service deployment for each workflow transition, and the random algorithm makes the selection in a random way. By utilizing two types of information about computational resources - static data, characterizing the speed of the processors, and dynamic data, showing the current load on the resources, provided by the Performance Monitoring and Analysis service, the performance-oriented algorithm chooses the deployments on the resources which provide the most appropriate execution conditions, with respect to performance. The prediction-oriented algorithm aims at optimizing the workflow execution time, using knowledge-based execution time estimations provided by the KAA. The algorithm is an extension of the Heterogeneous Earliest Finish Time (HEFT) algorithm which proved to be an efficient approach for scheduling of scientific workflows [2]. Finally, the prediction-and-performance-based algorithm is a combination of the two algorithms described above. It uses the knowledge-based performance predictions modified by the dynamic monitoring data, obtaining in that way reliable execution time estimations which are applied to optimize workflow execution time.

The Scheduler has been implemented as a web service, and integrated with the rest of the workflow orchestration and execution environment. It was tested for several pilot applications used in K-WfGrid, deployed on geographically distributed sites of the K-WfGrid testbed. One of the scheduling techniques (performance-based scheduling) was also successfully applied in the ASKALON project [3].

The Scheduler is an important part of the K-WfGrid environment. It covers the performance requirements of the knowledge-based project, making the use of the performance data provided by different system components. The main scientific innovation is to combine knowledge-based predictions with performance monitoring data, to support scheduling. The Scheduler provides also an innovative modification to the classical HEFT algorithm, originally designed for simple workflow DAGs (Directed Acyclic Graphs). The algorithm has been adjusted to the novel workflow model applied in K-WfGrid, based on colored Petri nets including different abstraction levels.

Acknowledgements. This research has been supported by the Project K-WfGrid (EU IST-2002-511385).

References
[1] K-WfGrid project - http://www.kwfgrid.eu
[2] M. Wieczorek, R. Prodan, T. Fahringer, Scheduling of Scientific Workflows in the ASKALON Grid Environment, ACM SIGMOD Record, 09/2005
[3] T. Fahringer, R. Prodan, R. Duan, F. Nerieri, S. Podlipnig, J. Qin, M. Siddiqui, H.-L. Truong, A. Villazón, M. Wieczorek. ASKALON: A Grid Application Development and Computing Environment. 6th International Workshop on Grid Computing, © IEEE Computer Society Press, November 2005, Seattle, USA
[4] M. Wieczorek, R. Prodan, T. Fahringer, Comparison of Workflow Scheduling Strategies on the Grid, In Proceedings of 6th International Conference on Parallel Processing and Applied Mathematics (PPAM 2005), LNCS 3911, © Springer-Verlag,
[5] D. Theiner and M. Wieczorek, Reduction of Calibration Time of Distributed Hydrological Models by Use of Grid Computing and Nonlinear Optimisation Algorithms, Proceedings of the 7th International Conference on Hydroinformatics (HIC 2006), September 2006, Nice, France.
[6] M. Wieczorek, R. Prodan, T. Fahringer, Dynamic Programming Based Approach for Bicriteria Workflow Scheduling on the Grid. Poster, HPDC (IEEE International Symposium on High Performance Distributed Compting) 2006, 19-23 June

8. DIPAS: Distributed Performance Analysis Services for Grid Workflows

Hong-Linh Truong (1), Peter Brunner (1), Thomas Fahringer (1), Francesco Nerieri (1), Robert Samborski (1), Bartosz Balis (2), Marian Bubak (2), Kuba Rozkwitalski (2)
(1) Institute of Computer Science, University of Innsbruck, Austria
(2) Institute of Computer Science, AGH, Poland

The performance analysis services in the K-WfGrid project aim at providing online information about the performance of Grid workflows as well as Grid resources involved in the workflow execution. Such performance information provides not only to the user insights into the execution of workflows but also to the K-WfGrid middleware and services, e.g., workflow execution and scheduler, knowledge about the performance for semi-automatically constructing and executing workflows.

In this poster, we describe DIPAS, a novel integrated tool which supports online performance monitoring and analysis of service-based workflows. DIPAS provides performance analysis and interpretation for workflows. That is performance analysis is based on a novel classification of workflow performance overheads. Performance constraints can be specified; based on that DIPAS informs the user and client performance problems by interpreting performance metrics at runtime. Performance analysis of workflows is integrated with that of Grid infrastructure into a single framework. Therefore, performance problems and faults can be easily detected and correlated by examining complex dependencies among Grid workflows and resources. Moreover, DIPAS provides a Web portal for the user to conduct performance monitoring and analysis of Grid workflows and infrastructure, thus substantially simplifying the way the user interacts with the performance tool.

DIPAS introduces several novelties by supporting performance monitoring, analysis, and interpretation of advanced workflows composed from Web/Grid services at multiple levels of abstraction and by unifying the analysis of Grid workflows and infrastructure into a single system.

Acknowledgments. The work described in this poster is supported by the European Union through the IST-2002-511385 project K-WfGrid.

References
[1] K-WfGrid Project website: http://www.kwfgrid.eu
[2] Hong-Linh Truong, Peter Brunner, Thomas Fahringer, Francesco Nerieri, Robert Samborski, Bartosz Balis, Marian Bubak, Kuba Rozkwitalski, "K-WfGrid Distributed Monitoring and Performance Analysis Services for Workflows in the Grid", July 2006. On submission.
[3] Hong-Linh Truong, Bartosz Balis, Marian Bubak, Jakub Dziwisz, Thomas Fahringer, Andreas Hoheisel, "Towards Distributed Monitoring and Performance Analysis Services in the K-WfGrid Project", In Proceedings of 6th International Conference on Parallel Processing and Applied Mathematics (PPAM'2005), LNCS 3911, (c) Springer-Verlag, Poznan, Poland, 11-14 September, 2005.

9. GEMINI - Generic Monitoring Infrastructure for Grid Systems and Applications

Bartosz Balis (1,2), Marian Bubak (1,2), Jakub Dziwisz (2), Kuba Rozkwitalski (2)
(1) Institute of Computer Science, AGH University of Science and Technology, Krakow, Poland
(2) Academic Computer Centre CYFRONET AGH, Krakow, Poland

The primary goal of the GEMINI is to provide a universal monitoring infrastructure able to work with virtually any monitoring tool, both for application and infrastructure monitoring. Existing monitoring tools can be adapted to GEMINI as plug-ins. GEMINI provides two web-service interfaces based on XML query languages: a monitoring interface based on PDQS (Performance Query Subscription) language, and an instrumentation interface based on WIRL (Workflow Instrumentation Request Language). This allows developers of tools that need to access monitoring data to focus on building their solutions, not on interaction with various tools. GEMINI can provide on-line access to monitoring data but this data is also stored in a persistent repository from which it can be collected later. All these features are very important for the purpose of performance analysis, scheduling and knowledge extraction in the K-Wf Grid project [1].

GEMINI consists of two main parts that are involved in data acquisition and transportation between the monitored entity and the client, namely the Sensor Infrastructure, and a network of Monitors. Sensors perform actual data extraction. A generic sensor library is provided to enable implementation and deployment of sensors in a standardized way. Monitors expose end-user interfaces and are responsible for resource dissemination and discovery, as well as hold performance data repository. Requests for data come to a Monitor, where they are parsed and analyzed to choose a proper data source - a sensor. The sensor is requested to actually provide the data which is then returned to the client [2].

An important achievement of GEMINI is a comprehensive support for instrumentation and monitoring of Grid workflows at different levels of abstraction - from the entire workflow to code regions of invoked applications [3,4]. GEMNI's instrumentation interface is used to control the instrumentation of applications, which is a necessary step before monitoring of applications is possible. GEMINI supports the concept of an intermediate representation of the monitored application, called SIRWF (Standard Intermediate Representation for Workflows). SIRWF allows the user to identify the parts of applications to be instrumented and specify the corresponding metrics to be evaluated, such as start/end events. Currently GEMINI provides tools to automatically instrument Java byte-code classes. After the instrumentation, a SIRWF description and probes are inserted into the code. Though this instrumentation is inserted statically, its execution is conditional and can be enabled or disabled at runtime. Consequently, the overhead of inactive instrumentation is negligible. GEMINI also handles, through adaptation of the OCM-G monitoring system, monitoring and dynamically enabled instrumentation of MPI applications written in C [5].

Current development comprises, among others, an integration of GEMINI with a peer-to-peer distributed-hash-table (P2P DHT) systems in order to support a fully decentralized architecture at the same time maintaining a fast resource dissemination and discovery [6].

The main scientific and technical achievements of GEMINI are the following [7]:
- Providing a framework for integration of diverse monitoring data sources into a common service-oriented infrastructure,
- Comprehensive support for monitoring of Grid workflows including dynamically-enabled fine-grained instrumentation that supports the concept of intermediate abstract representation of an application.
- Achieving interoperability while not losing performance of monitoring by combining service-based interfaces for querying and efficient RPC-based channels for streaming of data.
- Addressing security requirements with respect to data integrity and privacy with support for secure monitoring of resources behind firewalls or in private networks.

Acknowledgements. This research has been partially supported by the Project K-WfGrid (EU IST-2002-511385) and related SPUB-M grant. The authors thank the K-WfGrid Consortium for valuable comments.

References

[1] K-Wf Grid Project: http://www.kwfgrid.eu
[2] B. Balis, M. Bubak, J. Dziwisz, H.-L. Truong, and T. Fahringer. Integrated Monitoring Framework for Grid Infrastructure and Applications. In P. Cunningham and M. Cunningham, editors, Innovation and the Knowledge Economy. Issues, Applications, Case Studies, pp. 269-276, Ljubljana, Slovenia, October 2005, IOS Press
[3] B. Balis, M. Bubak, and K. Guzy. Fine-Grained Instrumentation and Monitoring of Legacy Application in a Service-Oriented Environment. In V.N. Alexandrow, G.D. van Albada, P.M.A. Sloot, and J. Dongarra, editors, Proc. 6th Intl Conference on Computational Science - ICCS'2006, vol. 3992 Lecture Notes in Computer Science, pp 542-548, Reading, UK, May 2006. Springer
[4] B. Balis, H.-L. Truong, M. Bubak, T. Fahringer, K. Guzy, and K. Rozkwitalski. An Instrumentation Infrastructure for Grid Workflow Applications. In R. Meersman and Z. Tari et al., editors, Proc. OTM 2006 Conferences, Lecture Notes in Computer Science, Montpellier, France, Nov 2006. Springer, accepted

[5] H. L. Truong, Bartosz Balis, M. Bubak, J. Dziwisz, T. Fahringer, and A. Hoheisel. Towards Distributed Monitoring and Performance Analysis Services in the K-WfGrid Project. In Proc. PPAM 2005 Conference, Lecture Notes in Computer Science vol 3911, pp. 157-163, Poznan, Poland, September 2005, Springer
[6] B. Balis, M. Bubak, J. Dziwisz, and K. Rozkwitalski. Towards a Generic P2P Monitoring Intrastructure for Grid Resources and Applications. In Proc. 2nd CoreGRID Workshop on GRID and Peer to Peer Systems Architecture, Paris, France, Jan 2006
[7] H.-L. Truong, P. Brunner, T. Fahringer, F. Nerieri, R. Samborski, B. Balis, M. Bubak and K. Rozkwitalski. K-WfGrid Distributed Monitoring and Performance Analysis Services for Workflows in the Grid, e-Science'2006, Amsterdam 4-6 Dec 2006, accepted

10. Grid Organizational Memory - Semantic Framework for Metadata Management in the Grid

Bartosz Kryza(1), Jan Pieczykolan(1), Marta Majewska(2), Renata Slota(2), Marian Babik(3), Adrian Toth(3), Jacek Kitowski(1,2), Ladislav Hluchy(3)
(1) ACC CYFRONET AGH, Krakow, Poland
(2) Institute of Computer Science, AGH University of Science and Technology, Krakow, Poland
(3) Insititue of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia

Grid Organizational Memory is the central information and knowledge source in the K-Wf Grid system [1]. It manages various kinds of metadata stored in the form of OWL ontologies.

The main functionality of the GOM knowledge base is oriented towards supporting the semi-automatic workflow composition from services available in the Grid [2]. This includes management of semantic descriptions of resources, data and services as well as ontological representation of knowledge about the domains to which the Grid is applied, which enable sufficiently rich semantic annotations of data and services. The managed ontologies can be stored in persistent storage and queried by standard query languages for RDF and OWL. Apart from the knowledge base itself GOM has several accompanying tools that support several typical use cases in the K-Wf Grid environment.

The architecture and design of Grid Organizational Memory make it a generic framework for knowledge management in the Grid [3]. The ontologies defining the Grid metadata can be divided into ontology components according to the ontology separation scheme [4]. Each ontology component can be managed by a separate instance of GOM Engine component, possibly distributed, which can be configured with such options as persistent store and some DL reasoner. This makes it possible to choose the best configuration for each ontology component. Each ontology component can be also published under proper URL thus giving clients easy access to the ontology. The ontologies managed by GOM can be updated by a special event system designed for fine grained modifications of OWL ontologies what makes integration of GOM with other system components easier combined with mechanism of subscription for change notifications.
A special web based component (called GOM Admin) has been developed to manage distributed deployments of GOM. It allows administrators to start, stop and monitor status of running instances of GOM Engine components. Several tools that support users in interacting with GOM have been developed. First one, called WSRF2OWL-S [5-6], has been developed to enable semi-automatic generation of semantic descriptions of services. This tool supports users by generating an OWL-S description of a service from its WSDL definition and a mapping from
WSDL to OWL which can be specified in a file or using a GUI integrated into K-Wf Grid GridSphere portal. Another tool is GOM Tab plugin for Protégé environment which enables manipulation of ontologies stored in GOM using Protege's advanced user interface and all its plugins for visualization of ontologies. A tool which translates from Common Information Model schema to OWL ontologies enables migration from legacy information systems. The prototype of the tool enables users to generate OWL ontology of CIM schema and convert instances from a CIM repository.

The main achievements of this work include the definition of unified semantic metadata model of the Grid called ontology separation scheme, design and development of generic distributed knowledge base with an event system enabling updating of managed ontologies and recovery of the state of the knowledge base from any given time in the past from serialized history of incoming events and development of Protege plug-in that enable easy interaction with the knowledge base.

The main scientific innovations include the architecture of flexible knowledge base framework, that can be optimally configured for a given kind of ontology, development of WSRF2OWL-S tool which enhances current state of the art by handling stateful Grid services and prototype implementation of CIM2OWL translation tool and development of special protocol for interacting with OWL ontologies, which allows the knowledge base to return a OWL document with a consise subset of the ontolog that contains a semantic neighborhood of results of a query.

Grid Organizational Memory has been evaluated within the K-Wf Grid platform by supporting both middleware level functionality such as infrastructure monitoring as well as application level scenarios [7-8], especially the ones involved in workflow composition process. Future work will include evaluation of different distribution models such as P2P, improvement of performance of the system as well as integration of ontology similarity algorithm for querying and updating the ontologies [9-10].

The paper presents detailed description of the Grid Organizational Memory architecture, a justification of its applicability in the context of K-Wf Grid project, and gives a thorough comparison of GOM with other existing semantic metadata solutions.

Acknowledgements. This research has been partially supported by the Project K-WfGrid (EU IST-2002-511385). The authors thank the K-WfGrid Consortium for many feedbacks during the design and development process.

References

[1] K-WfGrid Project - http://www.kwfgrid.eu

[2] K. Krawczyk, R. Slota, M. Majewska, B. Kryza, J. Kitowski: Grid Organization Memory for Knowledge Management for Grid Environment. In Bubak, M., Turala, M., Wiatr, K. (eds): Proceedings of the CGW'04, pp. 109-115, ACC CYFRONET AGH, Krakow, Poland, 2005

[3] B. Kryza, J. Pieczykolan, M. Babik, M. Majewska, R. Slota, L. Hluchy and J. Kitowski: Managing Semantic Metadata in K-Wf Grid with Grid Organizational Memory, In Bubak, M., Turala, M., Wiatr, K. (eds): Proceedings of the Cracow Grid Workshop '05, pp. 66-73, ACC CYFRONET AGH, Krakow, Poland, 2006

[4] B. Kryza, M. Majewska, R. Slota and J. Kitowski. Unifying Grid Metadata Representations Through Ontologies, In R.Wyrzykowski et.al. (eds) Proc. of 6-th Intl. Conf. on Parallel Processing and Applied Mathematics PPAM'2005, Poznan, Poland, R. Wyrzykowski et.al. (eds): LNCS 3911, Springer-Verlag, pp. 683-690, 006

[5] M. Babik, L. Hluchy, J. Kitowski and B. Kryza: WSRF2OWL-S: A Framework for Generating Semantic Descriptions of Web and Grid services. In Bubak, M., Turala, M., Wiatr, K. (eds): Proc. of the Cracow Grid Workshop '05, pp. 49-46 , ACC CYFRONET AGH, Krakow, Poland, 2006

[6] M. Babik, L. Hluchy, J. Kitowski, B. Kryza: Generating Semantic Descriptions of Web and Grid Services, to appear in Proc. of the 6th Austrian-Hungarian Workshop on Distributed and Parallel Systems, Springer, Innsbruck, Austria, 2006

[7] L. Dutka, B. Kryza, K. Krawczyk, M. Majewska, R. Slota, L. Hluchy, J. Kitowski: Component-expert Architecture for Supporting Grid Workflow Construction Based on Knowledge. In Cunningham P. Cunningham M. (Eds): Innovation and the Knowledge Economy. Issues, Applications, Case Studies. pp. 239-246, Vol. 2, IOS Press 2005

[8] M. Babik, O. Habala, L. Hluchy, M. Laclavik, M. Maliska: Semantic Grid Services in K-Wf Grid, to appear in Proc. of the 2nd International Conference on Semantics, Knowledge and Grid (SKG2006), IEEE Computer Society, Guilin, China, 2006

[9] J. Pieczykolan, B. Kryza, J.Kitowski: Semi-automatic Creation of Adapters for Legacy Application Migration, In proc. of Int. Conf. on Computational Science 2006, pp. 252-259, vol. 4, Springer, The University of Reading, UK, Springer, 2006

[10] B. Kryza, M. Majewska, J. Pieczykolan, R. Slota nad J. Kitowski. Grid Organizational Memory - Provision of a High-Level Grid Abstraction Layer Supported by Ontology Alignment. Future Generation Computer Systems, in print.

11. Peer-to-Peer Distribution Model for Grid Organizational Memory Knowledge Base

Bartosz Kryza(1), Jan Pieczykolan(1), Jakub Wach(2), Mikolaj Zuzek(2), Jacek Kitowski(1,2)
(1) Academic Computer Center CYFRONET AGH, Krakow, Poland
(2) Institute of Computer Science, AGH University of Science and Technology, Krakow, Poland

Latest trends in application of knowledge bases in distributed settings are being oriented toward scalable and fault-tolerant solutions based on Peer-to-Peer communication model. Currently, several semantic knowledge base solutions which are based on the mentioned model exist. However they are either very generic, with low performance and very basic functionality or are application specific oriented, making them awkward to use as semantic metadata management solutions for the Grid.

Grid Organizational Memory is a distributed knowledge base, developed within the framework of EU-IST K-Wf Grid project [1], for managing semantic metadata of Grid resources, such as hardware, data and software, as well as domain specific application information, using OWL-DL language as an ontology formalism. It is based on a notion of ontology separation scheme [2], where a global ontology is divided into interrelated ontology components, each managed by separate and possibly distributed GOM Engine element. A single ontology component is bound to a unique URI, and can be rendered and published as an OWL document. Each GOM Engine can be configured in different way, providing various capabilities for ontology storage and reasoning. Currently, GOM contains rather simple distribution model, assuming that each ontology component is treated as one whole and should be managed by a single GOM Engine element. In order to extend its functionality for very large metadata sets, authors propose a P2P model for distributing an ontology component into a network of peers that will together provide a scalable semantic metadata store for some part of the metadata, for example data registry.

Due to the nature of the Grid, the authors discuss that the unstructured model of a P2P network is more natural for such environment, since it keeps metadata close to the described resources. Also, the ontology separation scheme imposes grouping of peers by super-peer nodes, which manage the particular ontology component. This model assures natural clustering of metadata into consistent sets of ontological individuals, since metadata is stored 'locally' with respect to annotated resources. Unstructured P2P network model imposes less burden on administrators and users in terms of depoying new instances of ontology components. However, use of the mentioned model have some serious drawbacks, the most important is a high cost of query answering due to the need of flooding the network in order to retrieve all potential answers. This paper discusses possible solutions for that problem including distributed indexes such as DHT and subscription mechanism which should substantially decrease the network traffic necessary for answering most query patterns. Another issue discussed in the paper is the semantic reasoning over distributed ontologies. Although the assumption is that local models will be consistent, thus some reasoning over each local model will be possible, not always all possible entailments will be inferred in this way.
The applicability of several possible approaches to this problem is analysed. Since the proposed P2P model must also support the internal Grid Organizational Memory event system, which is used to update the semantic metadata on fine grained level (e.g. values of relations), a section is provided that shows how to extend that event system for a P2P network of nodes, assuring consistency of information.

The paper contains careful and thorough overview of current state of the art in the area, concentrating on semantic metadata stores for the Grid as well as on various new technologies that could help in designing this architecture and development of the prototype [3-4].

References

[1] K-Wf Grid Project - http://www.kwfgrid.eu
[2] B. Kryza, M. Majewska, R. Slota and J. Kitowski. Unifying Grid Metadata Representations Through Ontologies, In: Proc. of 6-th Intl. Conf. on Parallel Processing and Applied Mathematics PPAM'2005, R.Wyrzykowski et.al. eds., 2005, LNCS 3911, Springer-Verlag, pp. 683-690. Poznan, Poland.
[3] M. Cai, M. Frank, B. Pan, and R. MacGregor, "A Subscribable Peer-to-Peer RDF Repository for Distributed Metadata Management," Journal of Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 2, Issue 2, 2005.
[4] M. Castro, M. Costa, and A. Rowstron, "Debunking some myths about structured and unstructured overlays", NSDI'05, Boston, MA, USA, May 2005.

12. OnTal: A Tool for Ontology Extension in Knowledge-based Grid Systems

Renata Slota (1), Joanna Zieba (2), Maciej Janusz (1), Jacek Kitowski (1,2)
(1) Institute of Computer Science, AGH University of Science and Technology, Cracow, Poland
(2) ACK Cyfronet AGH, Cracow, Poland

The paper presents an approach to knowledge evolution developed within the EU IST K-Wf Grid project [1], applied to the process of Grid service registration. The project, which aims at enabling automated workflow composition, makes extensive use of various semantic resources gathered in distributed knowledge base called Grid Organizational Memory (GOM) [2]. GOM is divided into multiple ontologies which are either generic or domain-specific. Knowledge about Grid services registered in the system (stored in the form of OWL-S descriptions) is especially important, since as it is used in the process of workflow
composition and orchestration.

Extension of GOM ontologies can be useful in several cases. These include exploitation of knowledge that occurs in situations, which have not been predicted previously, thus omitted in the GOM repository (e.g. including informations from monitoring system, processing users' feedback and using external ontologies). The use-case of registering new services is particularly complex because both instances stored in the ontology and the ontology schema change. For facilitating this process the OnTal tool has been constructed and validated. The OnTal tool presented in this paper is designed to support this process. It uses ontology alignment methods to enable controlled, semi-automated extension of Grid Organizational Memory.

The paper describes the architecture and implementation issues of the tool, together with the ontology alignment methods and similarity measures adopted for the study. Interoperability of the tool with other K-Wf Grid modules is explained. The presented algorithm is also provided with examples of ontology extension from different pilot applications of K-Wf Grid. Moreover, the role of user interaction in knowledge evolution process is explained.

To sum up, the paper presents the ontology extension tool, its underlying alignment and similarity methods, and the process of registering new services while extending domain knowledge. Although developed within the K-Wf Grid framework, it is of general purpose and can be used independently.

References

[1] K-WfGrid project website, www.kwfgrid.net.
[2] B. Kryza, R. S?ota, M. Majewska, J. Pieczykolan, J. Kitowski, Grid Organizational Memory -- An Overview

of Semantic Approach to Providing High-Level Grid Abstraction Layer, FGCS, 2006, in print.

13. Knowledge Assimilation Agent - Component for Workflow-Based Grid Service Performance Prediction, Workflow Result Estimation and Semi-autonomic Service Discovery

Zoltan Balogh(1), Matrin Seleng(1), Martin Maliska(1), Ladislav Hluchy(1), Renata Slota(2), Joanna Zieba(2), Jacek Kitowski(2,3)
(1) Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
(2) Institute of Computer Science, AGH University of Science and Technology, Krakow, Poland
(3) Academic Computer Centre CYFRONET AGH, Krakow, Poland

Knowledge Assimilation Agent (KAA) is a knowledge-based component for Grid service workflows, which comprises three basic functionalities: assimilates run-time information about Grid services from different sources and produces performance estimations of future Grid service invocations (the KAA-Web Service); performs past workflow analysis and produces workflow result estimations (the KAA-WXA tool) and discovers new potential services through interactive semi-automatic ontology alignment (the OnTal tool). Individual tools responsible for each of three functionalities are described in the article. Use of KAA functionalities in scope of the K-Wf Grid 6th FP IST project is described.

Performance estimations about existing Grid Services is needed by several components in the K-Wf Grid [1-2] platform such as Automatic Application Builder (AAB), Scheduler and User Assistant Agent (UAA). Based on information provided by KAA, AAB selects most stable hosts out of the all available ones. AAB provides a list of hosts and KAA ranks individual hosts according to their stability (how often they failed in the past). Historical information about services are used. The results are then passed from AAB to Scheduler, which makes use of such information. Scheduler is another consumer of KAA estimations. Scheduler needs to select which deployment of a service is the best to invoke in a given context. Estimation of service performance is dependent on the invocation context, i.e. what are the invocation parameters and input resources of a service. KAA needs to know the expected invocation context before making the estimation. UAA is the next component using knowledge assimilated by KAA. Workflows from the past are compared with the current workflow using KAA-WXA. If identical workflow is found, the result of such workflows is offered to the current user for potential reuse through UAA interface or user can used this composed and fully filled workflow with data to submit it for computing. OnTal, the tool for ontology evolution, follows one of general ideas of K-WfGrid, that broadening the knowledge of the system, stored in the form of semantic resources, allows more effective automatic workflow composition (performed by such components as WCT and AAB). Moreover, as the tool is semi-automatic, it is helpful for users who have to update the ontologies during services' registration.

The central part of KAA is the KAA-Web Service (KAA-WS) [3], which is a stateless web service implementation through which third party components request and retrieve the generated knowledge. KAA-WS implements the core learning algorithm, concretely an extended and improved instance-based learning technique [3 - 6]. There is an internal ontology model of Grid service performances, resources and used invocation parameters which is called Knowledge Assimilation Schema (KAS) [7-10]. KAA consumes information generated by third party components such as Gemini (a monitoring infrastructure) and the Workflow Composition Tool (WCT) and transform them into KAS representation. New information are added upon completion of any grid workflow, which is detected by a component called KAA-WCT. Using a retrieved workflow identifier, a component called KAA-Gemini retrieves information about Grid services invocations and transforms them into KAS representation. KAA-WXA checks the GWES XML database for new workflows, gets the context from the workflow (based on XPath technique) and sends it to the UAA in a note format. If UAA has the same note (same context and input data) the old note is overwritten by the new one so users have always actual information about new results. To extend knowledge gathered in the system during the process of new Grid services' registration which is a part of service discovery the OnTal tool has been designed [11-12], see also [13].
It processes a OWL-S service description, maps its external ontological concepts onto concepts already existing in the system knowledge base and proposes a set of ontology extensions validated further by a knowledge administrator.

The main achievements of this work include ability of estimation of Grid service performance measures (run-time, availability, accessibility, etc.) based on the invocation parameters and input resource meta data and semi-automatic extension of the knowledge base by ontology evolution methods.

The main scientific innovations of KAA are extension of instance based learning (IBL) technique of WS performance prediction by enabling retrieval and inclusion of semantic resource properties into the feature vector, enabling specification of WS performance prediction profiles which specify feature vector and result to be estimated and proposition of a novel ontology for IBL-based WS performance prediction together with methodology extending the classical Case-Based Reasoning. Development of knowledge extension by ontology evolution using a combination of lexical and structured similarity and considering neighbourhoods of two ontology concepts when their similarity is assessed is another achievement of KAA in the context of workflow construction for grid applications.

Acknowledgements. This research has been partially supported by the Project K-WfGrid (EU IST-2002-511385). The authors thank the K-WfGrid Consortium for many feedback during the design and development process.

References

[1] K-WfGrid Project - http://www.kwfgrid.eu
[2] HABALA O., BABÍK M., HLUCHÝ L., LACLAVÍK M., BALOGH Z.: Semantic Tools for Workflow Construction In: Proc.of International Conference on Computational Science, Part III, LNCS 3993, Springer-Verlag, 2006, pp. 980-987, ISSN 0302-9743, ISBN 3-540-34383-0, May 28-31, Reading, UK.
[3] BALOGH Z., GATIAL E., LACLAVIK M., MALISKA M., HLUCHY L.: Knowledge-based Runtime Prediction of Stateful Web Services for Optimal Workflow Construction. Proc. of 6-th Intl. Conf. on Parallel Processing and Applied Mathematics PPAM'2005, R.Wyrzykowski et.al. eds., 2006, LNCS 3911, Springer-Verlag, pp. 599-607, ISSN 0302-9743, ISBN 3-540-34141-2. Poznan, Poland.
[4] BALOGH Z., GATIAL E., LACLAVÍK M., HLUCHÝ L.: Refinement of case retrieval for instance-based Grid service performance prediction through semantic description of input data. In: J.Paralic, J.Dvorsky, M.Kratky (Eds.): Znalosti 2006, Proceedings of 5th annual conference. VŠB-Techická universita Ostrava, Fakulta elektrotechniky a informatiky, 2006, pp.151-158. February 2006, Hradec Kralove, Czech Republic. ISBN 80-248-1001-8.
[5] BALOGH Z., HLUCHÝ L.: Knowledge Assimilation for Performance Prediction of Grid Services for optimal Workflow Execution. In: Proc. of 1-st workshop Grid Computing for Complex Problems - GCCP 2005, VEDA, 2006, pp.135-144, ISBN 80-969202-1-9. November-December 2005, Bratislava, Slovakia.
[6] BALOGH Z., GATIAL E., LACLAVIK M., HLUCHY L..: Refinement of Case Retrieval for Grid Service Performance Prediction through Semantic Description of Input Data. In: Bubak, M., Turala, M., Wiatr, K. (editors): Proceedings of the Cracow Grid Workshop '05, Cracow, November 2005, pp. 59-65, published by ACC CYFRONET AGH, Poland, April 2006. ISBN 83-915141-5-3.
[7] NGUYEN G., BALOGH Z., LACLAVIK M., GATIAL E., HABALA O., SIMO B., HLUCHY L.: Modeling and presentation generation for ontology-based knowledge in the experience management system. Proc. of 8th Intl. Conf. informatics, SSAKI Bratislava, 2005, pp. 77-81. ISBN 80-969-243-3-8.
[8] BALOGH Z., GATIAL E., LACLAVIK M., MALISKA M., HLUCHY L., ORAVEC V., NGUYEN G.: Predicting Web Service Performance based on Web Service States and Input Data Properties. Proc. of 8th Intl. Conf. informatics, SSAKI Bratislava, 2005, pp. 113-117. ISBN 80-969-243-3-8.
[9] BALOGH Z., GATIAL E., LACLAVÍK M., NGUYEN G.T., HLUCHÝ L.: Capture, Discovery and Reuse of Knowledge in REMARK. In: ICETA 2004. 3rd International Conference on Emerging Telecommunications Technologies and Applications. Košice ELFA, 2004, pp.89-93. ISBN 80-89066-85-2.
[10] LACLAVIK M., SELENG M., GATIAL E.,BALOGH Z.,HLUCHY L.: Ontology based Text Annotation – OnTeA; In: Proc. of 16-th European-Japanese Conference on Information Modelling and Knowledge Bases, EJC'2006, Y.Kiyoki et.al. eds., 2006, Dept.of Computer Science, VSB - Technical University of Ostrava, pp. 280-284, ISBN 80-248-1023-9. Trojanovice, Czech Republic.
[11] ZIEBA, J., SLOTA, R., MAJEWSKA, M., KRYZA, B., and KITOWSKI, J.,: Ontology Alignment and Ontology Similarity for Extension of Service Ontology in Grid Environment. In: Bubak, M., Turala, M., Wiatr, K. (eds.): Proceedings of the Cracow Grid Workshop '05, Cracow, November 2005, published by ACC CYFRONET AGH, Poland, April 2006. ISBN 83-915141-5-3.
[12] KRYZA, B., SLOTA, R., MAJEWSKA, M., PIECZYKOLAN, J., KITOWSKI, J., Grid Organizational Memory -- An Overview of Semantic Approach to Providing High-Level Grid Abstraction Layer, FGCS, 2006, in print.
[13] SLOTA, R., ZIEBA, J., JANUSZ, M., KITOWSKI, J., OnTal – the tool for ontology extension in knowledge-based grid systems, CGW'06.

14. Service-based Flood Forecasting Simulation Cascade in K-Wf Grid

Ondrej Habala, Martin Mališka, Ladislav Hluchý
Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia

Flood Forecasting Simulation Cascade is a flood prediction application, using series of meteorological, hydrological, and hydraulic simulation models to predict possible flood risks. This application has been developed for several years, and in the K-Wf Grid project has been implemented as a set of grid services, using the WSRF standard. In this article we describe its architecture, implementation details, usage scenarios, user interfaces, and our experiences in using WSRF standard and K-Wf Grid middleware for flood prediction.

Prediction of natural disasters is an important part of every public early warning and risk assessment system. This applies also in the area of flood prediction. The process of flood prediction is of complex nature, since it requires knowledge of future weather conditions, which is a complex requirement in itself. The Flood Forecasting Simulation Cascade (FFSC) application of the K-Wf Grid EU IST project is a system, which attempts such complex task, using a set of several simulations, performed in stages - weather prediction first, followed by watershed integration, hydrological prediction for the target river, and finally hydraulic simulation for the target area. The application also contains several visualization modules, which can be used to display results from all stages in several ways - for example, the final water flow simulation can be displayed as a series of 2D images, or as a 3D animation, suitable also for 3D display systems, like the CAVE environment. Each stage of the application can be performed by multiple simulation models, depending on the characteristics of the target area and on user preferences.

The application has been previously developed using older Globus 2 middleware and executed as a set of binary modules in grid in the 5th Framework Programme, in the scope of the CROSSGRID EU IST project. The application is distributed in nature and has many possible scenarios, which make it an ideal target for the SOA-oriented grid, composed of web and grid services, using WSDL and WSRF standards. We have extended the application with additional simulation models and visualization tools, and implemented all these as WSRF-based services, deployable in the Globus 4 environment. The application is composed of 19 different service interfaces, each with multiple instances, and more different services are ready to be developed and incorporated into the application. This loosely-coupled architecture allows us not only to add new simulation models, but to create several cooperating administrative domains. This is very important for the area of flood management - usually, each river in a country is managed by different river authority, and rivers spanning multiple countries are managed by different authority in each country. These organizations are independent, but have to be able to cooperate and to exchange data. SOA-based FFSC allows just this by exposing secure service interfaces for simulation execution, data retrieval and visualization.

One of main targets of the K-Wf Grid project is improved usability of grid middleware. FFSC is one of several case studies for this target. The application is exposed through the K-Wf Grid's workflow management system and its generic grid access tools, but also through a set of customized, domain-oriented user interfaces, which are easy to use for domain experts, yet conformant to the overall K-Wf Grid architecture. This allows users to harness the power of distributed grid infrastructure, without the need to be grid experts, or even skilled IT users. All data inputs is provided through application-specific input forms, and results of the application are displayed in visualized form, so users don't have to access grid tools for data transfer virtually at all. Important fact is that all these interfaces are part of the K-Wf Grid architecture, so any application can be integrated into K-Wf Grid middleware in such manner.

Since K-Wf Grid middleware relies heavily on knowledge stored in an ontological structure, FFSC has been also described in this manner. All services have been annotated using the ontology schema of K-Wf Grid, and are easily discoverable for the K-Wf Grid workflow management engine.

The Flood Forecasting Simulation Cascade is important case study not only for K-Wf Grid, but as a grid application in general. It is a complex distributed and resource-demanding application, it leverages the computational and data resources of K-Wf Grid testbed, yet it is accessible to users with virtually no grid experience. It can be easily extended with more simulation models and other service interfaces, and it naturally supports cross-domain collaboration of several independent administrative entities.

Acknowledgements. This research has been partially supported by the Project K-WfGrid (EU IST-2002-511385). The authors thank the K-WfGrid Consortium for many feedback during the design and development process.

References

[1] K-WfGrid Project - http://www.kwfgrid.eu

[2] HABALA O., BABÍK M., HLUCHÝ L., LACLAVÍK M., BALOGH Z.: Semantic Tools for Workflow Construction In: Proc.of International Conference on Computational Science, Part III, LNCS 3993, Springer-Verlag, 2006, pp. 980-987, ISSN 0302-9743, ISBN 3-540-34383-0, May 28-31, Reading, UK.

[3] NGUYEN G., BALOGH Z., LACLAVIK M., GATIAL E., HABALA O., SIMO B., HLUCHY L.: Modeling and presentation generation for ontology-based knowledge in the experience management system. Proc. of 8th Intl. Conf. informatics, SSAKI Bratislava, 2005, pp. 77-81. ISBN 80-969-243-3-8.

[4] HLUCHY L., HABALA O., NGUYEN G., SIMO B., TRAN V., BABIK M.: Grid Computing and Experience Management in EU RTD Projects of IISAS. In: Proc. of 1-st workshop Grid Computing for Complex Problems - GCCP 2005, VEDA, 2006, pp.7-19, ISBN 80-969202-1-9. Nov-Dec 2005, Bratislava, Slovakia

[5] HABALA O., CIGLAN M., HLUCHÝ L.: Data management in Flood Prediction. In: Advances in Grid Computing - EGC 2005, Proc. of 3-rd European Grid Conference, LNCS 3470, Springer-Verlag, 2005, pp. 872-879, ISSN 0302-9743, ISBN 978-3-540-26918-2, Feb 2005, Amsterdam, The Netherlands

[6] HABALA O., CIGLAN M., HLUCHÝ L.: Data Management in the FloodGrid Application. In: Bubak, M., Turala, M., Wiatr, K. (editors): Proceedings of the Cracow Grid Workshop '04, Cracow, December 2004, pp. 221-228, published by CYFRONET AGH, Poland, April 2005. ISBN 83-915141-4-5.

15. Application of K-WF Grid Technology to Coordinated Traffic Management

Stefano Bianchi, Marco Masetti, Gianni Viano
Softeco Sismat S.p.A., Genoa, Italy

The Coordinated Traffic Management (CTM) pilot application developed by Softeco Sismat [1] over the K-WF Grid middleware [2] targets the computation of the emission of traffic air pollutants in a urban area and has been developed in tight collaboration with the Urban Mobility Department of the Municipality of Genoa, which provided a monolithic implementation of the model for the pollutant emission calculations, the urban topology network and real urban traffic data. The objective is computational-wise challenging and provided a valid testbed for the K-Wf Grid middleware, allowing to evaluate and demonstrate the benefits of the application of the project results to a real application environment as well as to deliver pre-competitive application frameworks and tools to be exploited in concrete business opportunities. As required by the internal architecture of the K-Wf grid middleware, the CTM application workflow has been divided into several different steps in order to allow the semi-automatic composition of services and the definition of a set of ontologies which describe the CTM domain and feed the system with the information needed for the proper selection and execution of services.

Traffic management algorithms apply on a city network. City network can be seen from two different perspectives: a topological view (districts polygonal, nodes coordinates, distances, areas,...) and a graph view (basically a DAG graph). The main CTM application functionalities currently implemented, deployed and tested are:
-- Best route calculations: given two different city districts first topological calculations are applied to compute the district starting and ending internal nodes, then for each starting node the Dijkstra single source shortest path algorithm is applied;
-- Traffic flows calculations (over the best routes, given real traffic data): main traffic flows are derived from an O/D (origin/destination) matrix: flows are then split per path taking into consideration paths length and vehicles everage speed on path links;
-- Air pollutant emission calculations: an existing model (PROGRESS - PROGramme for Road vehicles EmiSSions evaluation in Genoa, developed by the University of Genoa) has been ported on the grid middleware. The model can take care of a massive amount of data and handles up to 8 different vehicles categories. It computes the emissions for the 4 most critical pollutants (CO, HC, NOx e PM).
-- Data graphical representation: graphical services are used to plot computations results in diagrams and charts. Analysis results are also represented in SVG (Scalable Vector Graphic) format for an easy access from any web browser.

Each functionality can be considered as a self-consistent scenario and can be used singularly or as an intermediate step for more complex use cases. In order to enrich the test bed of the pilot application and to offer the possibility of evaluating some benchmarks, for each aforementioned functionality implementations with different payloads have been registered into the middleware knowledge base and made available to the user.

With respect to a traditional market solution, the CTM application exploits all the advantages that come from the adoption of a grid of resources instead of a classical monolithic solution:
-- parallel and seamless task execution: the middleware is able to run independent tasks in parallel and on different grid nodes whenever possible, exploiting the power of the grid;
-- computations potentially dispersed on different grid nodes;
-- dynamic task execution depending on resource monitoring: performance monitoring and feedback information is used to constantly tune task execution for a better use of grid resources.

Furthermore, the application benefits from the adoption of the K-WF Grid middleware, raising then from common grid applications:
-- Semi-automatic application workflow construction: using the domain, service and data application knowledge, the middleware is able to compose the best workflow that fits the user needs;
-- A new application development model: provided all the services to cope with each application task, the application workflow is built by the middleware; the user can anyway decide to save and reuse already composed workflows and re-compose them at some events (new services deployed, changes in the grid of resources, new scenarios). A derived advantage with respect to monolithic market solutions is the possibility to cope with new scenarios that can arise in the traffic management domain;
-- Services potentially automatically exploited in different domains: the more services are described in the middleware knowledge base, the more scenarios are reachable, as they can be reused independently and self-consistently.

The introduction of the aforementioned benefits in a complex domain such as the CTM application can be appreciated considering a real scenario, for example the estimation of the consequences of closing a downtown road on the daily traffic and on the air pollution. A conventional approach in fact would usually consist in performing manually several simulations in order to estimate the consequences of different scenarios and control measures. An experienced traffic manager would edit the inputs, calibrate the parameters, launch the simulation, collect the results, perform the analysis, come to conclusions and plan actions. All these activities would be usually performed sequentially, on a single PC, with manual data management and offline evaluations - a waste of time and resources. K-Wf Grid fosters and supports the transition from a work organisation modelled on workflow based on “humans” which perform tasks by means of a number of monolithic applications, toward a “virtual workflow” implemented by a middleware with the transparent orchestration of grid services.

The CTM application has been a valid test bed for the K-WF Grid middleware and helped to cope with and solve middleware weaknesses that arose in the first stages of application development.
The application has been helpful to test core middleware functionalities and the robustness of the overall system. Currently CTM services are deployed on grid nodes at Softeco Sismat and University of Innsbruck premises. As a result of the collaboration between an industrial IT partner and a potential real end user, the application also allowed an early evaluation of the impact of the adoption of grid-based and K-Wf Grid-powered solutions in a domain which actually offers interesting business possibilities.

This research has been partially supported by the Project K-WfGrid (EU IST-2002-511385). The authors thank the Mobility Department of the Municipality of Genova for the support in domain analysis and application conception, design and evaluation.

[1] Softeco Sismat SPA - http://www.softeco.it
[2] K-WfGrid Project - http://www.kwfgrid.eu

16. Enterprise Resource Planning Pilot in K-WFGrid

Pantelopoulos Stilianos, Kalaboukas Konstantinos
Singular SA, Greece

Most enterprises are under mounting pressure to adopt and deploy reliable, high performance Enterprise Resource Planning (ERP) solutions. Such enterprises are expected to find ways to enhance their competitiveness by offering new or improved services a) to their employees, who need better and faster access to business information; and b) to their customers who demand faster response to their requirements and improved quality of service and support.
Moreover, from the technological point of view, Service Oriented Architecture (SOA) is the de-facto technological basis for implementing state of the art ERP platforms. Grid technology leverages SOA by grid enabling services. However, the combination of Grid technology and SOA raises some issues that KWF-Grid aims to undertake.

As current enterprise environment is characterized by rapid changes and fuzzy networks of inter-enterprise correlations and dependencies, the effective decision making, which constitutes crucial factors concerning the company positioning and competitiveness in the enterprise environment, requires computation- and data-intensive tasks due to the complexity of the supporting algorithms and the massive storage of enterprise data. In addition, the enterprise employees, who are the real ERP users, demand fast access to machine processed enterprise data so as to support efficiently and execute effectively back-office business processes, leading to the faster delivery of vital information, to the optimization of the enterprise performance and to the enhancement of customers' satisfaction.
To solve computation- and data- intensive tasks and to integrate back-office business processes, K-WfGrid provides a reliable, horizontally scalable platform for dynamic, semi-automated composition and optimized execution of business processes, delivering and ensuring high ERP and database throughput and performance at service levels.
Taking into consideration the business lifecycle and the everyday activities of an enterprise some ERP usage scenarios have been identified regarding the competitiveness and the total performance of the enterprise. The "Products Stock Management" scenario has been the most critical one, concerning the calculation of the required quantities for each product type, traded by the enterprise, to be ordered in a weekly / monthly basis, so as the future products stock level in the enterprise warehouses to remain above a predefined, safe level. The "Products Stock Management" scenario utilizes several time-series analysis and prediction models (i.e. both the Single and the Triple Exponential Smoothing models [1]), to predict the forthcoming sales, and calculates the next-period orders of each product type traded by the enterprise, based on the long-term product sales and stock management algorithms. This module is practically a module of the Data Warehouse management component of the ERP.
Indicatively a FMCG* enterprise trades over than five thousand different product types, and places over ten orders per product type and per month to its suppliers, which makes over fifty thousand orders per month.
K-WfGrid provides the ERP user (e.g. Accountants, Sales and Finance Managers who are responsible for the pricing policy and the roder placement), who has no experience in grid infrastructures and workflow management systems, with an integrated tool and user friendly interface to execute parameterized, scenario-specific, optimized business processes.
The adoption and deployment of K-WfGrid in enterprise environments, characterized by complex intra- and inter- organizational business processes that orchestrate business-driven services deployed by customizable business software systems, scale existing systems horizontally using a grid-enabled service-oriented

architecture. More specifically, K-WfGrid approach and platform:

- Allows semantically-assisted search, discovery and selection of appropriate tasks/services, stored in services repository that could typically contain some hundreds of services, with their desired functionality in order to compose a business-driven e-workflow and to establish connections among these tasks (control and data flow);
- Allows the user to identify, at design time, the operational metrics of discovered services and grid resources, including timeliness, quality of products delivered, cost of service, and reliability, facilitating, thus, the composition of optimized grid-enabled e-workflows;
- Supports the semi-automated generation, storage, and reuse of business e-workflows with a level of parameterization;
- Shortens calculation time of data- and computation- intensive business-driven ERP-based processes, and improves enterprise competitiveness delivering instantly high quality computations; and
- Enables simple ERP users, with no experience in grid computing and workflow management, to benefit from the emerging Grid-enabled Service Oriented Computing by offering user-friendly interfaces to a complex Grid-based environment.

References

[1] National Institute of Standards and Technology (NIST) / SEMATECH, e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 21st Aug 2006

----

* Fast Moving Consumer Goods (FMCG) industry encompasses a huge range of products and services in manufacturing, distribution and retailing from food, beverage & tobacco, toiletries, cosmetics & packaging, to consumer health & household goods.

17. Software Engineering Aspects on K-WfGrid

Marian Bubak, Maciej Malawski, Piotr Nowakowski, Steffen Unger

An important yet often neglected aspect of European and other international research projects is the proper application of software engineering methodologies, relevant for such large-scale distributed scientific collaborations. This paper aims to present some of the methods involved in successful management and control of software engineering in a Grid development project, on the example of Knowledge Workflow Grid (K-WfGrid) which is a FP6 IST undertaking.

K-WfGrid, as a Specific Targeted Research Project, aims to develop Grid sofware components for semiautomatic generation of advanced application workflows, deployment of these workflows and reuse of information gathered during execution to make informed decisions at the next stage of workflow creation.

Due to the scope and structure of the project a custom software engineering approach has been worked out by the Project consortium, basing on accepted software development methodologies [1] but also taking into account the specifics of developing software within an international collaboration and their impact on such processes as software design, prototyping, testing and quality assurance.

Software engineering in K-WfGrid covers the following aspects:
- K-WfGrid quality objectives and their role in software development,
- organizational structure of the project, including project Work Packages (WPs), project tasks (the basic organizational unit of the project) and project management,
- project documentation expressed in the form of deliverables and publications produced by K-WfGrid partners. This concerns both the physical layout of K-WfGrid deliverables and the deliverable submission and acceptance procedure, involving the Internal Review Board and Quality Assurance. An official K-WfGrid Publication Policy is also defined,
- generic conventions, standards and metrics which the Project adheres to, including:
  - tools used within K-WfGrid (for publication as well as programming),
  - coding conventions for all programming languages used within K-WfGrid,
  - the Central Repository and its means of usage,
  - the release preparation process,
  - other standardized conventions to be followed.
- the K-WfGrid software development process and the role of QA procedures at each step of the Project s timeline. This involves both the technical aspects of the project as well as the reporting that goes with it and is divided into the following sections:
  - the software requirements review (initial phase),
  - the software design review (first development phase),
  - unit testing procedures (all development phases),
  - release and prototype reviews (all development phases),
  - testbed QA (all development phases),
  - publication reviews (all phases),

- managerial reviews (all phases).
- the K-WfGrid corporate identity as a set of rules, captions and graphical elements to be used in official Project publications.

The paper describes in detail the quality assurance methodology and guiding rules together with processes supporting their realization in all phases of the software development process, from user requirements analysis to integration testing. K-WfGrid software engineering methodologies have been set forth in the the K-WfGrid Quality Assurance plan, which bases on both accepted standards and previous experience with drafting integrated QA procedures for CrossGrid (such as the IEEE-compliant SRS templates, Design Specifications and the deliverable review procedure), and on the experiences of other affiliated European Grid projects, most notably the DataGrid Project [2] and GridLab [3]. IEEE template 730-1989 (see [4]) was used for the creation of the K-WfGrid QAP.

ACKNOWLEDGEMENTS

This research has been partially supported by the K-WfGrid project(EU IST-2002-511385). The authors thank the K-WfGrid Consortium for feedback received during the design and development process.

REFERENCES
[1] Eric J. Braude, Software Engineering: An Object-Oriented Perspective (John Wiley & Sons, Inc., 2001)
[2] The DataGrid Project Management Website; http://eu-datagrid.web.cern.ch/eu-datagrid/WP12/default.htm
[3] The GridLab Project Quality Assurnce Plan, available at http://www.gridlab.org/Project/Deliverables.html
[4] The IEEE Software Engineering Standards catalog; http://standards.ieee.org/catalog/olis/arch_se.html

18. GScript Editor as a Part of the ViroLab Presentation Layer

Wlodzimierz Funika, Piotr Pegiel
Institute of Computer Science AGH, Kraków, Poland

A correctly designed user interface is a very important task in many software projects. The user interface is a software layer that lays at the bottom of a stack - a user interacts with an application by using this layer. Therefore it is crucial for the user interface to provide the whole system functionality in a user friendly way. This paper presents an approach used in one part of the user interface from the ViroLab project.

The presentation layer in ViroLab project consists of three different parts. The first layer is the functionality that could be accessed from a web browser - it is designed to use web technology - especially portals e.g. GridSphere. The second layer is a simple command line tool. The last part is a tool that provides the script editor and many extra features within. This editor has to be very extendible and helpful for the user - e.g. with support for the code completion. This layer can also integrate a preprocessor connected to the code editor, semantic code inspection or ontology browser. It is also possible to connect the editor to the ViroLab Grid Environment, execute a script code and trace the execution.

In this paper we provide an overview of three possible solutions for the extensible environment for the script editor (GScript). Two of them are based on Eclipse framework, while the third one is a standalone application based on functionality provided by jEdit - an open source text editor written in java.

The first approach is to create an editor as a plug-in for the Eclipse framework. Eclipse provides many functionalities that can be adopted in an easy way to the script editor - like code completion and syntax coloring. Many developers have got used to the Eclipse style of the presentation - so they do not have to spend a long time learning a new environment (component orientation and positioning are managed by Eclipse in a common way). One of the greatest problems is the size of the application built into Eclipse - it is over 100 MB.

Another solution is to build an editor using Eclipse RCP approach. The RCP is a part of the Eclipse project created to simplify the process of developing Rich Client Applications. The advantages are the same as in the previously presented solution. But by using RCP we can eliminate many disadvantages - e.g. the size of the application that we have created for the test purpose with Eclipse core components was about 10 MB.

The last solution is to create standalone application. We can base it on the jEdit, which has many features, such as easiness to create an editor with syntax highlighting, or auto indenting etc. But it does not support syntax completion and many extra features like ontology browser or grid resources browser have to be implemented from scratch.

References
1. Eclipse C/C++ Development Tooling - CDT - http://www.eclipse.org/cdt/

2. Eclipsepedia RCP description - http://wiki.eclipse.org/index.php/Rich_Client_Platform
3. JEdit home page - http://www.jedit.org/
4. EclipseCon 2006 - Rich Client Application Development by Jean-Michael Lemieux - http://wiki.eclipse.org/images/d/d9/EclipseCon_RCP_Tutorial_2006.pdf

19. Virtual Laboratory in ViroLab

Marian Bubak (1,2), Tomasz Gubala (2), Maciej Malawski (1), Marek Kasztelnik (2), Tomasz Bartynski (1,2), Piotr Nowakowski (2)
(1) Institute of Computer Science, AGH University of Science and Technology, Krakow, Poland
(2) ACC CYFRONET AGH, Krakow, Poland

The main objective of the ViroLab project is developing a virtual laboratory for infectious diseases that facilitates medical knowledge discovery and decision support for e.g. HIV drug resistance [1,2].

We have identified 3 groups of the virtual laboratory users: users of the ViroLab decision support system who use a dedicated web GUI to seamlessly access ViroLab applications (BAC, Rule Miner); experiment developers who plan ViroLab experiments using specific development tools; and experiment users who use prepared experiments to gather scientific results which may be shared with others using collaboration tools. Additionally, provenance of results should be tracked and recorded for future usage.

At the very beginning, when planning an experiment, the user does not know exactly what resources are available in the environment. Therefore we propose a language that allows users to express the planned processing by means of abstract Grid operations [3].
We describe the concept of experiment planning and the proposed architecture of the ViroLab virtual laboratory: presentation layer, collaboration tools, experiment planning system, runtime system, access to data, access to computing resources and provenance.

Experiments planned by users will be built from basic blocks. In order not to constrain the development of these elements to one technology, the virtual laboratory enables access to various types of resources (stateful Web Services, distributed components, batch jobs). Accessing different types of resources should not increase the complexity of experiments. For this reason we propose uniform access to the underlying middleware layer using an abstraction over it. The user does not have to posses knowledge on how to invoke specific types of resources. We have identified the following requirements for the middleware system: ability to handle invocation of each operation which forms a single step of an experiment in VL and to execute it on Grid resources; providing a standardized interface for operation execution; transferring input and output parameters using the Data Access layer; transparently selecting the best operation realization and interfacing with resource brokers; integration with existing infrastructures the Project partners have access to (EGEE and DEISA).

Acknowledgements.
This research has been supported by the Project ViroLab (EU IST STREP Project 027446). The authors thank the ViroLab Consortium for valuable feedback during the design phase.

References
[1] ViroLab - EU IST STREP Project 027446; www.virolab.org
[2] P. Sloot, C. Baucher, M. Bubak, A. Hoekstra, P. Paszczak, A. Posthumus, D. van de Vijver, S. Wesner, A. Tirado-Ramos: VIROLAB a Virtual Laboratory for Decision Support in Viral Diseases Treatment. Proceedings of Cracow Grid Workshop - CGW'05, November 20-23 2005, ACC-Cyfronet AGH, 2006, Kraków, pp. 33.
[3] T. Gubala and M. Bubak: GridSpace - Semantic Programming Environment for the Grid, In R.Wyrzykowski et.al. (eds) Proc. of 6-th Intl. Conf. on Parallel Processing and Applied Mathematics PPAM'2005, Poznan, Poland, LNCS 3911, Springer-Verlag, pp. 172-179, 2006

20. Management and Access of Biomedical Data in a Grid Environment

Matthias Assel, Bettina Krammer and Aenne Loehden
HLRS - High Performance Computing Center of Stuttgart, Germany

Cross-organizational data exchange is a daily routine that most people do not actually perceive. Sending e-mails attached with sensitive information to business partners is as normal as shopping on the Internet. In the case of directly accessing data at the appropriate location or sharing data across different partners, many companies are afraid of having their data abused not only by a third party, but in the same way even by a trusted organization.

Since an increasing number of applications - ranging from physics, chemistry, and aerospace to healthcare - require data at a very large scale, of both size and distribution, providing data over a grid across different organizations is becoming more and more popular. Access to these data resources must be carefully controlled by a sophisticated management system.
The complexity of data management on a grid arises from the scale, dynamism, autonomy, and distribution of such resources. To conceal these complexities of the underlying infra-structure, the middleware system has to ensure that the resources appear transparent to their users. This could be achieved by hiding the different data resources and their internals behind a layer of virtualization services that guarantees data access in a consistent, data resource-independent way.
In the context of biomedical data, the requirements for the virtualization services are of particular importance and differ from other grid applications. Besides standard mechanisms and interfaces for discovering and accessing data resources, for instance using Web Service standards, the transformation of heterogeneous data is, due to different formats of each resource, one of the most challenging issues. Therefore, the querying and handling of metadata plays a decisive role to obtain interoperability between data sets at different centers across different countries.
When dealing with confidential data, security mechanisms are of the utmost significance and are seen as the most critical part. Authentication and authorization are a precondition for transferring data over a grid, but the high sensitivity of biomedical information and personal data also demands secure transmission and secure storage. In order to keep the privacy and protect the confidentiality of patients, medical data need to be encrypted and even anonymized before using it in a virtual laboratory.

In this paper, a brief overview of existing solutions, like OGSA-DAI [1], EGEE gLite Data Management [2], and the SDSC Storage Resource Broker [3] shall be presented and evaluated, and one possible example of a virtualization layer, which allows secure access to biomedical data resources, will be proposed.

References:
[1] The OGSA-DAI Project. http://www.ogsadai.org.uk
[2] gLite - Lightweight Middleware for Grid Computing. http://glite.web.cern.ch/glite
[3] The SDSC Storage Resource Broker (SRB). http://www.sdsc.edu/srb/index.php

21. Enabling Commercial Chemical Software on EGEE Grid - Gaussian VO

Mariusz Sterzel, Tomasz Szepieniec
Academic Computer Centre CYFRONET, Academy of Science and Technology, Krakow, Poland

Today's computing and storage requirements make EGEE Grid a perfect platform for chemistry applications [1]. Despite the fact that a few quantum chemical codes have been successfully ported to the Grid [2] there is still lack of commercial chemical software that would be available for whole EGEE community. The main reasons blocking wide use of commercial software are its license requirements. Confirmation of fulfilling the license requirements is the key step in porting software to the Grid and than using it.

There are many computational packages available nowadays for chemists. Among them the most popular for its easiness of use, large number of computational methods and constant development during the last thirty years is the Gaussian package [3,4]. The advantages of Gaussian make it an excellent choice for our application [5]. Moreover, enabling to use the Gaussian package on the Grid by wider community will benefit not only in chemistry but also in biochemistry, medicine and even physics.

In this report we would like to share our experience in enabling commercial software on the Grid using Gaussian program [6] as an example. In such a case access to the software is influenced by additional license requirements. Therefore, in order to meet these requirements, a new virtual organization (VO) gaussian has been created. Membership in gaussian VO enables user an access to Gaussian package operated and maintained by ACC CYFRONET AGH. The access to Gaussian package suite is open for each EGEE Grid user who does have a valid certificate and agrees to obey Gaussian, Inc. license requirements.

Presented installation procedure may be used to make available other commercial software for EGEE community.

References;
1. EGEE web page http://www.eu-egee.org
2. GAMESS package on EGEE Grid: http://egee.grid.cyfronet.pl/gamess
3. W. J. Henre, W. A. Lathan, R. Ditchfield, M. D. Newton, J. A. Pople, Gaussian 70, QCPE, 11, 236,1973
4. Gaussian web page: http://www.gaussian.com
5. A. Danel, S. Zapotoczny, M. Sterzel and M. Pilch, Journal of Molecular Structure: THEOCHEM, 682, 179, 2004
6. Gaussian program on EGEE Grid: http://egee.grid.cyfronet.pl/gaussian

22. The Quest for Pharmacology Active 'Never Born Proteins' within EUChinaGRID Project

Irena Roterman-Konieczna(1), Marek Kochanczyk(1), Maciej Malawski(2), Monika Piwowar(1), Tomasz Szepieniec(3)
1) Department of Bioinformatics & Telemedicine Collegium Medicum UJ, Krakow, Poland
2) Institute of Computer Science AGH, Krakow, Poland
3) ACK Cyfronet AGH, Krakow, Poland

Contemporary pharmacology in it's quest for more relevant and effective drugs need to examine large range of biological structures to identify biologically active compounds. Even for relatively small polypeptide chain (less than 80 amino chains) a computation of three-dimensional structure of the molecule and determining it's function takes more than 40 minutes on average single CPU machine available today.

In our experiment that is realized in the frame of EUChinaGrid Project [1], the search for new biologically active compounds is focused on peptide-like molecules containing about 60 amino acids in polypeptide chains. The 'never born proteins' are planned to be constructed and examined in amount of about 10^7. The limited number of proteins existing in the nature will be extended to those, which have not be recognized in any organisms. The assumption is that those which do not exist in the nature may also demonstrate the biological activity, which directed on pharmacological use may correct some pathological phenomena. The time required to compute such number of sequences are so large, that we need to mobilize the several thousands CPUs to face the task in reasonable time. We consider large grid environments as the only platform to realize such computational projects.

For this purposes, the EUChinaGrid Project aims is to integrate the largest European grid (currently maintained by EGEE Project [3]) and similar Chinese initiative. The described application is planed to exploit resources from both infrastructures. Moreover, the integration within the project is not limited to the infrastructure only, but also enable collaboration of three teams that are working on presented subject: the authors of this paper are responsible for protein structure prediction basing on very promising, newly introduced model [2]; RomaTre University team generates the sequences and it is doing the same job as our team, although on the basis on different methods, which gives opportunity to verify the results; finally the Chinese team of experimentalists, are ready to synthesize these proteins, which appear in silico to be pharmacologically active.

[1] IST EUChinaGrid Project, web page: http://www.euchinagrid.org/
[2] M. Brylinski, L. Konieczny, I. Roterman. Fuzzy-Oil-Drop Hydrophobic Force Field - A Model to Represent Late-stage Folding (In Silico) of Lysozyme. Journal of Biomolecular Structure & Dynamics, Vol.23, Nr 5 (2006).
[3] IST EGEE Project: http://www.eu-egee.org/

23. A Flexible Communication Layer for Grid-based Visualization with VTK

Paul Heinzlreiter and Jens Volkert
GUP, Joh. Kepler University Linz, Austria/Europe

Following the evolvement of grid computing into an important tool for research, the visualization of large datasets produced on the grid has become increasingly important. In this context the visualization itself becomes a suitable grid application.
This especially refers to scenarios where the data to be visualized is already generated in a distributed manner or if a scientist wants to investigate the results of a simulation being executed on a remote supercomputer on his desktop machine.
For implementing parts of our grid-based visualization solution we have chosen the Visualization Toolkit (VTK) since it already offers a set of widely used visualization algorithms and can easily be extended to add the functionality required.
VTK already offers classes for point-to point communication over BSD sockets and parallel implementations of visualization algorithms based on the Message Passing Interface (MPI) standard.
One possible approach to execute VTK-based parallel visualization algorithms on the grid is to use MPI as communication layer together with a grid-enabled version of MPI such as MPICH-G2. However this limits the

flexibility regarding selection of resources and transmission modes as such low-level configurations are typically handled within the MPI implementation or the resource broker. Furthermore not all grid resources offer MPI implementations and the VTK library needs to be recompiled with the local MPI library to make use of it.
Consequently we extended VTK by adding point-to-point communication over Globus XIO similar to the socket interface and implementing wrapper classes to map the MPI communication model on top of multiple XIO connections. This enables the direct execution of parallel VTK visualization algorithms while using different resources in a coherent way by building on top of Globus.
Using Globus XIO the data communication over the grid can be configured separately for each connection according to the needs of the application as well as the underlying grid infrastructure. For example, GSI encryption can be used if there are security concerns for a specific connection. Additionally we have implemented an optional automatic compression method for transmitting large datasets. Meanwhile the VTK communication layer takes care of data management issues such as endianess conversion between resources with different architectures. Future work will build on top of this communication infrastructure while focusing on the construction and efficient execution of distributed visualizations on the grid.

24. Automatic and On Demand Web-Based Configuration-Driven Remote Build System

Francesco Giacomini(1), Roberto Lops(1), Elisabetta Ronchieri(1), and Francesco Prelz(2)
(1) INFN CNAF, Bologna, Italy
(2) INFN Milano, Italy

In a large project, software build is a difficult process when people are located in different sites and involved in the development of several software components subject to frequent modifications and enhancements. Therefore, it is important to setup the local daily build system in order to allow Release Managers to perform the monitoring and management of the whole project, and Testers to verify the latest functionalities and changes in the code. The system described supports storage and publishing of results on web pages, automatic e-mail notifications in case of problems (e.g., build failures), the remote build, artifacts and a user-friendly setup. These features are enabled by means of few configuration scripts. In this paper, we detail the overall structure of the daily build system used by the Release Managers and Testers at the INFN National Centre for Telematics and Computer Science.
Software development process has to be equipped with reliable builds of the software in order to reduce daily problems (e.g., build breaks). This is often not provided with our surprise. In addition, software development can be geographically distributed to several sites. People in the working group involved in this process can typically have a number of roles: (1) Site Managers, involved in the set up of the working group testbed, (2) Release Managers, involved in the build system in order to monitor the software and provide their packages, (3) Testers, responsible to verify and validate the software functionalities and (4) Developers, who write software. When the code size and complexity grow, the traditional "nightly" build process for the entire project can become inefficient and unmanageable. However, if individual sites are active on a well-defined code partition, local site-specific build system infrastructures can come into play, to improve the overall efficiency of the build process. When build failures occur, people who are responsible to maintain this system (the Release Managers), need to write and test patches as in collaboration with the relevant developers. If a problem cannot be corrected by the Release Manager, the developers are contacted. Consequently, in a large software development project, the ability of meeting deadlines relies on the effective communication between Release Managers and Developers, leading to prompt correction and detection of the build. A well-structured web-based access to information about the build process, is a useful mechanism to effectively support communication. Other important figures in geographically distributed projects are Testers who are in charge of verifying the new coming functionalities and the latest changes in order to support the software integration process. As Release Managers, Testers are in contact with the developers in order to quickly apply fixes to the bugs discovered during the testing activity. The build system used by a group of developers collaborating in the middleware Joint Research Activity (JRA1) of the Enabling Grids for E-sciencE (EGEE) project. The results and other useful information about the build system, such as the documentation related to the installation of specific services and to the software dependencies, are reachable from a web portal at link http://goldrake.cnaf.infn.it:8080/ibrido/home.php. The proposed build system supports the Release Managers' activity that consists of building the software and publishing packages used for testing. In fact, the Release Managers have to be able to build both vanilla software (i.e., no modifications, no customizations) and software that contains changes and new dependencies. In order to verify the build and quickly provide packages, it is important to tailor the local build system by handling remote build executions not only automatically but also on demand, and by publishing build information in an easy and understandable format.
The proposed system improves the Release Manager's productivity, and simplifies the daily job of Testers, by allowing to select either the latest packages, or just the ones belonging to a specific application.
Several requirements have been gathered to lead the definition of the system architecture. The system needs: 1. to be automated to simplify the Release Managers' work; 2. to be efficient to reduce Release Managers, Testers and Developers frustration and its abandon; 3. to be general enough in order to be applicable to a wide range of version control systems, software projects such as EGEE, LCG and other open-

source entities, build mechanisms such as autotool, ant; 4. to be easy to maintain to simplify its future improvements; 5. to provide fast feedback about results via Web pages and e-mail notification; 6. to be flexible to use for site administrators and the major roles in the development team; 7. to be able to handle complex dependency patterns.
The capability of handling very complex patterns of inter-module dependency is of paramount importance. Dependencies are fundamental in the build process as they affect the build structure by requiring that destination modules are built prior to the source. Dependencies are located at two different levels of the build system: 1. package such as project libraries 2. system such as system libraries.

25. A new Approach to Support Component Applications on Grid

Maciej Malawski (1), Tomek Bartynski (2), Eryk Ciepiela(1), Joanna Kocot(1), Przemyslaw Pelczar(1), Marian Bubak(1,2)
(1) Institute of Computer Science, AGH Univerity of Science and Technology, Krakow, Poland
(2) ACC CYFRONET AGH, Krakow, Poland

Running component based applications on the Grid environment is still a difficult and inconvenient. First, resources should be discovered across various administrative domains. Then, components must be deployed and properly connected [1].

We propose a solution which consits of a well defined application description language (ADL) and a tool that will take over the responsibility for launching the application.

MOCCAccino provides a novel ADL specification and a manager tool that makes deployment process transparent for the user. A base component environment is MOCCA [2], a CCA compliant framework that supports deploying components on remote resources and provides communication layer. It is build upon H2O, which is a secure and lightweight middleware platform providing a flexible component container [1].

ADLM (ADL for MOCCAccino) is a XML based language which allows users to specify their requirements, such as number of components and connections between them. Moreover, it enables to provide optional information that determines the resource on which component will be placed. ADLM enables user to describe application components whether they are computation or communication intensive.

Our prototype tool parses an ADLM file, retrieves static information about available resources, chooses the best location for components, deploys them and launches the MOCCA application. The system has three tiers:
- users machine on which client program is run,
- machine with H2O kernel that runs manager components,
- other machines with H2O kernels, which are registered as available resources.

It is possible to run components from the first two tiers on a single client machine.

The objectives of current our work are:
- support for dynamic changes in the resource pool,
- optimisation of the component collocation algorithms,
- test on a real problems applications and on large computing infrastructure.

Acknowledgements. This research was supported in part by the EU IST-2002-004265 Network of Excellence CoreGRID and the related SPUB-M grant.

References
[1] G. Stuer, V. Sunderam, J. Broeckhove. Towards OGSA Compatibility in the H2O Metacomputing Framework. Future Generation Computer Systems 21 (2005) 221-226
[2] Maciej Malawski, Dawid Kurzyniec, and Vaidy Sunderam. MOCCA towards a distributed CCA framework for metacomputing. In Proceedings of the 10th International Workshop on High-Level Parallel Programming Models and Supportive Environments (HIPS2005), 2005, IEEE

26. An Approach for Intra-VO Differentiation of Computing Services in Grid Systems

Sergio Andreozzi, Marco Cecchi, Vincenzo Ciaschini, Andrea Ferraro, Francesco Giacomini, Antonia Ghiselli, Alessandro Italiano, Gian Luca Rubini, Davide Salomoni
INFN-CNAF, Bologna, Italy

In current Grid production systems, users belonging to the same Virtual Organization (VO) have typically the same opportunity to access the available resources on a First-In-First-Out (FIFO) approach. This is a limit in the efficient sharing of resources, especially when VO's have a huge number of members with different tasks and use cases. In this paper, we describe a proposal for differentiating the access to computing resources among users of the same VO (this approach is described in the context of the gLite middleware).

Our approach grounds on a rigorous definition of service classes in terms of characterizing attributes that describe different quality of service levels other than the best effort. Examples of parameters characterizing a service class are the target share in the utilization of the site resources, policies related to the maximum walltime of a single job or a priority level. The set of service classes defined by a VO is called service model. According to the needs of each VO, resource providers will configure at the physical level the service classes to access computing resources relying on the local resource management system capabilities. The important constraint is that jobs submitted to the same computing facility under the same service class must follow a FIFO approach. As initial design, we envision three main service classes: best effort, guaranteed and express. For each service class, we describe the set of mandatory and optional parameters.

Once the service classes are defined and configured at each resource provider, users part of the same VO, but with different credentials regarding their group or role membership can be assigned to different service classes, thus enabling a differentiation of the access to resources of a VO.

Given the geographical distribution of a Grid, the high number of sites and resources, and the dynamics of usage of the system, it is essential for a VO to have mechanisms to dynamically change how the VO users are assigned to the different service classes. In order to achieve this goal, we enrich our model with a distributed policy framework such as G-PBox (it relies on XACML for policy definition). By means of this component, the VO manager can define mapping policies that assign a set of Grid credentials to a certain set of instances of service classes. When a job is submitted to a site, the authentication and authorization layer queries the distributed policy framework by passing the Grid credentials of the user (Grid identity and VO membership information). The authorization framework will answer with the service class to which the user should be mapped, if any.

The proposed approach is currently under testing in the INFN-Grid and relies on key middleware components like VOMS, G-PBox and LCMAPS. The proposed system enables a flexible intra-VO computing resources differentiation. By means of this approach, a VO can enforce that different groups of users have a guaranteed share of resources and the assignment can be dynamically changed over time without intervention at the sites.

27. On Translating Common Information Model to OWL Ontology

Marta Majewska (1), Bartosz Kryza (2), Jacek Kitowski (1,2)
(1) Institute of Computer Science, AGH University of Science and Technology, Kraków, Poland
(2) Academic Computer Centre CYFRONET AGH, Kraków, Poland

The need of unified semantic description of the various grid aspects appeared during the work on the knowledge based system for composing workflows for the grid environment in the EU IST K-Wf Grid project [1], where OWL ontologies were chosen for metadata descriptions of the Grid.
The set of ontologies describing generic and domain specific features of the grid were gathered and integrated for the thematic areas of workflows, grid applications, services, data and resources. In particular, according to the workflow centered nature of the project and for its chosen pilot applications dedicated workflow and application ontologies were developed. OWL-S [2] was used for the services description and the ontology for grid resource description was an open issue. Among reference ontologies for modeling the hardware and software computer resources the DMTF Common Information Model (CIM) [3] as well known, organizationally supported and regularly updated ontology for mentioned area, (e.g. popularly referred in software for management of systems, networks, users and applications across multiple vendor environments) seemed promising. The CIM ontology is being composed of the core and common models, with set of extensions provided for software or hardware by producers and users. CIM is based on a native representation language called Managed Object Format (MOF) [4]. As OWL was a representation choice for the project ontologies the problem of the interoperability appeared. To keep the unified manner of the description of the grid and to enable knowledge management and acquisition functionality for the external components demanding ontology consistency checking and the reasoning functionality, from many solutions for providing ontology interoperability a translation from CIM to OWL was chosen. However, the reconciliation of the two formalisms originating from the various backgrounds was challenging. CIM inherited features from both object oriented modeling and database modeling. Contrary, OWL came from the field of artificial intelligence, peculiarly from among description logic based formalisms. Mappings between two ontological representations were proposed and are discussed in the paper in detail. The achieved ontology is presented. The paper contains also thorough comparison of existing approaches to conversion of CIM to OWL.
The tool is described that performs the conversion of CIM schema and allows converting CIM instances to OWL individuals. A use case scenario is presented, showing benefits of expressing CIM system descriptions in OWL.

References:
1. K-Wf Grid project website, http://www.kwfgrid.net/.
2. OWL-S: Semantic Markup for Web Services, http://www.w3.org/Submission/OWL-S/.
3. DMTF, Common Information Model (CIM) Standards, http://www.dmtf.org/standards/cim/.
4. CIM Managed Object Format (MOF), http://www.dmtf.org/education/mof/

28. Quality of Service on the Grid with User Level Scheduling

Jakub Moscicki (1,2), Marian Bubak (3,2), Peter M.A. Sloot (2)
(1) CERN - IT, Geneve, Switzerland
(2) Informatics Institute, Universiteit van Amsterdam, The Netherlands
(3) Institute of Computer Science AGH, Krakow, Poland

Currently the largest Grids lack an appropriate level of the Quality of Service (QoS) in two ways: the infrastructure and middleware is not enough reliable and a simple, batch-oriented processing model is suboptimal for a number of applications. User-level scheduling is a light software technique that enables new capabilities to be added and QoS characteristics and reliability to be improved, on top of the existing Grid middleware and infrastructure.

User-level scheduling techniques may be used to reduce the job turnaround time and to provide a more stable and predictable job output rate. Splitting the processing into many fine-grained tasks improves the load balancing and ensures that the workers are used efficiently. As the result the computing resources may be returned to the Grid faster. We discuss the implications of this technique for the users, the application developers and the resource providers.

Applications which have been interfaced with the user-level scheduler include High Energy Physics data analysis, Monte Carlo simulation, Biomed applications and others. Distributed frequency analysis and autodoc-based drug discovery are the recent large-scale activities and are summarized below.

In May and June, CERN successfully supported a series of large-scale data-processing activities carried out by the International International Telecommunications Union (ITU) as part of the ITU's Regional Radiocommunication Conference. Several sites of the EGEE infrastructure provided a computing Grid of more than 400 PCs to work on each analysis in parallel, and the processing was conducted using the user scheduling layer. The system completed more than 200 000 very-short frequency analysis jobs (clustered in around 40 000 processing tasks) in around one hour, proving that on-demand computing with a short deadline is possible on the Grid.

Earlier this year the same technique was used to perform a sizeable fraction of an in silico drug discovery application using the EGEE and other Grid infrastructures. The challenge was to analyze possible drug components against the avian flu virus H5N1. This activity showed that a user-level scheduler can improve the distribution efficiency on the Grid from below 40% to above 80% by optimizing the allocation of the fine-grained computing tasks. Efficient automatic-error recovery mechanisms proved to be efficient in extended periods of continuous work (30 days).

References
[1] C. Germain-Renaud, C.Loomis, J.T.Moscicki, R.Texier: "Scheduling for Responsive Grids" to appear in: 2006, Journal of Grid Computing (special issue on EGEE User Forum)
[2] J.T.Moscicki: "Distributed Geant4 Simulatrion in Medical and Space Science Applications using DIANE framework and the Grid", Nuclear Physics B, 125 (2003) 327-331
[3] Web links: http://cern.ch/diane, http://cern.ch/twiki//bin/view/ArdaGrid/WebHome

29. A Failover Solution for EGEE-compliant stateful Grid Core Services

Lukasz Flis, Jakub Janczak, Marcin Radecki,
ACC CYFRONET AGH, Krakow, Poland

Grid environments like EGEE [1] attracts constantly growing number of scientists by making available computations on a scale not possible before. The scientific community requires a certain level of the infrastructure availability to carry out the experiments smoothly. In current EGEE-compliant architecture the users may access the resources via so called core services that are essential for operation of grid information system, computational resources and data management tools. This approach facilitates the usage of grid but makes the availability of core services a crucial issue.

Within Central European Regional Operation Centre we developed a solution for improving the availability of a Resource Broker (RB) [2] - the service responsible for accepting grid jobs from users and submiting them on grid as well as further control. Although there can be many Resource Brokers over the grid however, a job sent through a particular Resource Broker instance can be controlled using only the same machine. Such approach can cause an inconvenience to users when the machine fails. Inability to retrieve results on time or control the job is often the main reason to resend lost jobs through different Resource Broker service while previous set of jobs is still running consuming resources uselessly.

To increase availability of the service and get rid of downtimes related to hardware and software failures, we have introduced some portion of redundancy which removed set of potential point of failures. Although, due to the Resource Broker is a stateful service it requires us to maintain the data synchronization between the machines.

Existing failover solution [3] present similar approach, but are dependent on external file storage services. Such file services needs to be reliable and highly available what implies use of two or more additional machines and reliable shared storage device. However, the advantage of this approach is encrypted network improving data security. Our solution seems to be easier to deploy and more cost effective by reducing amount of required hardware. Fully functional setup can be started by adding only one ordinary machine to the existing installation positioning our solution in a low-cost area.

We base our approach on opensource software [4], [5] plus small fraction of own RB-specific modifications. Our goal was to create a mechanism which can be easily deployed and integrated with current grid installations. We also kept in mind to limit modifications to current core services middleware. Currently, our mechanisms are being tested in production grid giving promising results.

The main achievement of that work is a relatively low-cost solution that brings critical grid services availability to the higher level. Service downtime related to middleware updates, hardware maintenances and failures is significantly reduced. Process of service migration is transparent for service clients and in most cases unnoticeable to the grid users. Resources are migrated automatically when node related failure is detected. Tests being performed in production grid at ACC CYFRONET AGH are showing that the mechanism integrates smoothly with current grid middleware, it is resistant to new grid middleware upgrades. Aggregating both open source software and low-cost solution advantages, it can be easily deployed at any grid center running core services.

It seems that adding some service-related modifications will allow the failover mechanisms to be applicable to other critical grid core services. In future work we are going to analyze the possibility to adapt our solution to other stateful services.

The paper presents the motivation for our work, design goals, implementation and overview of deployment procedure followed by results of evaluation in EGEE production grid.

[1] EU EGEE project web page http://www.eu-egee.org/
[2] EGEE project Resource Broker development web page
http://glite.web.cern.ch/glite/packages/R3.0/deployment/lcg-RB/lcg-RB.asp
[3] A. Kiryanov, High Availability of Core LCG Services, 2005
http://cern.ch/kiryanov/LCG-HA\ Documentation.pdf
[4] The High-Availability Linux Project: www.linux-ha.org
[5] Network replicated block device - DRBD: http://www.drbd.org/

30. A General Approach for the Highly Pluggable Workflow-processing Environment of the Next Generation UNICORE

Henning Mersch, Bernd Schuller, Achim Streit
Forschungszentrum Juelich, Germany

Workflow processing in distributed systems and Grids includes a huge range of activities today. Running complex applications in a service-oriented environment requires both low-level service orchestration and a domain-oriented way to define the high-level application workflow. Consequently, it is next to impossible to choose a single workflow description language that will fit all the needs. We expect many workflow representations to coexist in a typical "next-generation" Grid. Many workflow description languages already exist, such as BPEL, that was originally intended for web service orchestration. Furthermore, many graphical user interfaces for creating and monitoring workflows have been developed in the Grid and eScience communities. See for example Taverna as well as in the context of business process management (like jBPM).

In this work, we present a flexible, generic system for workflow processing that is fairly independent of the possible description languages and workflow processing engines that are used. The only assumption made is, that the workflow is generated in some form of a graph having vertices and edges. As will be shown, the use of plugin points for both vertices and edges assures that the proposed system is widely applicable.

This work is motivated by the need for a versatile end-user client for the next generation of the UNICORE Grid system. Existing UNICORE client software is focused on providing access to low-level services, suitable for "Grid expert" users. However, in a lot of relevant use cases, the users are non-Grid experts or even non-Computer-Scientists, who will be daunted by such a client.

31. Synchronizer - Based Dynamic Workflow Implementation in Grids

J. Borkowski (1), D. Kopañski (1), M. Tudruj (1,2)
(1) Polish-Japanese Institute of Information Technology, Warsaw, Poland
(2) Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

This paper shows how to organize workflow execution control based on predicates computed on internal global states of activities. We discuss Grid-level parallel program control mechanism based on application global states monitoring. At the cluster level, application processes report their local states to special control processes called synchronizers. Synchronizers construct consistent application states and compute predicates on them. Depending on the predicates, control decisions are taken. The decisions are distributed among processes in the form of control signals. The signals can activate control actions associated with them in processes and influence application behaviour in asynchronous manner. They can change the effective contents of application processes and modify the order in which they are executed. This control method de-couples the control aspects solved using a high-level abstraction, from the computational code. This approach has been embedded in the PS-GRADE graphical program design environment [1], which is an extension of the original P-GRADE system, Cluster level program execution control by synchronizers has been extended towards co-ordination of applications executed on a Grid [2]. PS-Grade cluster applications constitute component applications of large Grid-level applications. They are controlled by additional Grid-level synchronizers, which collect state reports from applications, construct Grid-level global states, evaluate predicates on them and send control signals to applications.

The paper describes dynamic workflow control implementation based on synchronizers. It proposes generalized design principles, which extend basic workflow control paradigms, since the control is based on programmable predicates computed on workflow execution states. This control method enables convenient workflow programming from elements that show run-time modifiable functionality and dynamic interactions. An overview of dynamic workflow control patterns [3] implemented with synchronizers is presented.

References

[1] M. Tudruj, J. Borkowski, D. Kopanski, Graphical Design of Parallel Programs with Control Based on Global Application States Using an Extended P-GRADE System, in Distributed and Parallel Systems, Cluster and GRID Comp., Kluver, Vol. 777, 2004.
[2] D. Kopanski, J. Borkowski, M. Tudruj, Co-ordination of Parallel Grid Applications Using Synchronizers, Proceedings of PARELEC 2004, Dresden, Germany, IEEE, pp 323-327.
[3] W.M.P. Van der Aalst et al., Workflow Patterns, Distributed and Parallel Databases, 14, (1), pp. 5-51.

32. Charon Extension Layer - Universal Toolkit for Grid Applications and Computational Jobs Maintenance

J. Kmunicek (1), M. Petrek (1,2) and P. Kulhanek (2)
(1) CESNET z. s. p. o., Praha, Czech Republic
(2) National Centre for Biomolecular Research, Brno, Czech Republic

This contribution describes Charon Extension Layer (CEL) system - a universal framework creating a layer upon the basic Grid middleware environment and making an access to the complex Grid infrastructure much easier compared to the native middleware. The CEL system unifies the variability in Grid middleware (PBS, LCG/gLite, etc.) allowing transparent access to distinct Grids. Moreover, the CEL system offers easy access and utilization of heterogeneous Grids in a unique, easy and smoothly integrated way.

Charon Extension Layer toolkit provides a command-line oriented interface and is supposed for users that require a full control over their running computational jobs. CEL system provides uniform and modular approach for (complex) computational jobs submission and management and forms a generic system for the use of application programs in the Grid environment independently of Grid middleware present at specific fabric infrastructure. CEL can be easily used for powerful application management enabling single/parallel execution of computational jobs without the job script modification. Simultaneously, standard job management involving easy job submission, monitoring, and result retrieval can be performed without any additional hassle or requirements on users.

Here we present a new implementation of the high-level middleware core services provided by CEL. Specific parts of the system were rewritten to enable new set of features to be available for end users utilization. First of all, present version of CEL is capable to provide multi-Grid approach (through so-called "sites") that makes everyday Charon utilization much more comfortable. Moreover, several further enhancements as inclusion of different data transfer synchronization modes, deeper personalization of the user's Charon environment, guided handling of grid middleware error states, module name completion, and others are included too. Currently CEL system is successfully used for end users research work within the application-generic Virtual Organization for Central Europe (VOCE) Grid infrastructure, part of the EGEE II Grid.

33. Mechanisms for Java Application Adaptive Resource Management in NGG

Arkadiusz Janik, Krzysztof Zielinski
Department of Computer Science, AGH-UST, Cracow, Poland

Next Generation Grids (NGG) is principally concerned with the architecture for a rich, open infrastructure to support a very wide range of applications. An implied NGG requirement is that it be possible to build and maintain applications with predictable levels of performance and reliability, resilient to changes in the underlying infrastructure, even for large distributed and long-running applications. Incremental evolution must be supported so that software revisions can be managed without loss of service. This will be based on an integrated approach to management of infrastructure and applications based on policies, with no assumption of centralised control or global knowledge of system-wide state.

Service Level Agreement is central to the conceptual model of NGG provided business services and therefore forms a key aspect of the underlying Grid infrastructure. A service-level agreement defines the nature and consequences of an interaction between a service provider and a consumer. It could be transformed into policies of Grid applications management. Enforcement of the policy-driven management requires system support providing monitoring, decision point, and effectors implementation on different virtualization level of Grid system architecture.

The proposed in this paper solution referrers to Grid system virtualized as a cluster of MVMs (Multitasking JVM) and construction mechanisms supporting policy-based resources management. For this environment we present Transparent Resource Management (TRM) architecture that can be used as a system-level solution that provides mechanism for Java Application Resource Management in Next Generation Grids. Features of TRM system suit most of NGG requirements such as resource on demand, self-management, dynamic resource management etc. The end-user of TRM system is able to run Java application with arbitrary chosen adaptability strategy. A strategy is defined as a finite state machine. Every state defines policy of managing one or more resources. The policy is defined as a set of constraints for each resource. The strategy also defines conditions of transition before states. The application user is able to set constraints on how the application is processed. The strategy is applied during application deployment. Presence of adaptability strategy makes that application runs in a policy-driven way. The current policy of resource management that is applied to an application is chosen by the autonomic manager in a transparent way outside the application.

In this paper we explain details of the architecture of TRM system and show how it can be used to build NGG solutions that guarantee requested level of SLA by better control of grid resources sharing and their proper virtualization. The presented examples referring to Java applications running over MVMs show that system can be used to control allocation of application's resources depending on application requirements. It is also very good illustration of the fact that a quality of the obtained solution is very much depended on system mechanisms provided by underlining virtualization technology.

The paper also shows how proposed architecture can be used in NGG to build the system where every group of users in VGRIDs has predefined set of policies that can be applied to applications run by user to meet this user's privileges and constraints.

34. Magrathea -- Grid Management Using Virtual Machines

J. Denemark (1,3), M. Ruda (2,3), L. Matyska (2,3)
(1) Faculty of Informatics, Masaryk University, Brno, Czech Republic
(2) Institute of Computer Science, Masaryk University, Brno, Czech Republic
(3) Cesnet z. s. p. o., Praha, Czech Republic

With varying requirements of grid users and virtual organizations they belong to on computing resources, their configuration, and quality of service, site administrators have to decide which configuration should be supported by which computing nodes, which nodes should be dedicated to specific groups of users, etc.

Obviously, this may lead to serious waste of computing power and unnecessary limitations on number of available nodes which meet specific requirements. Using virtualization techniques, each physical node can provide several logical nodes with different operating systems and various configurations at the same time. Jobs running in one logical node may be easily preempted by jobs with higher priority running in different logical node, which results in better utilization of resources.

In this paper, we contemplate deployment of virtual machine (VM) technology on computational clusters which are used by various grid user communities. Usage scenarios include, for example, preemptive job execution, starting job in a requested operating system with specific software configuration, checkpointing a job etc. We also provide a description of a prototype implementation of integrating PBS with virtual machines that serve as a basis for integrating MetaCentrum resources into EGEE infrastructure.

Currently, there are several open source virtualization solutions available, for example, Xen, OpenVZ, VServer. While Xen creates a complete virtual hardware on top of which standard operating system may run, the other two solutions share the same kernel for all virtual machines and can be seen as an enhanced change root environment with better isolation and resource restriction. All these approaches are discussed and compared in the paper. We also compare our approach with the workplaces ued in a Globus environment.

Using virtual machines on worker nodes also brings several problems. For example, cluster scheduling system must be modified to be able to correctly recognize available resources (the logical nodes). Allowing smart preemptive job execution may require suspending of running jobs which is all but trivial for parallel tasks and results in the need for sophisticated scheduling algorithms. Monitoring and benchmarking tools may be confused as it is not possible to use all advertised computing resources at the same time.

Although, for simplicity, we chose Xen as an underlying virtualization technology, the results concerning techniques which have to be enhanced and areas requiring more research are quite general and can also be applied to the other approaches.

To verify the possibility to deploy virtual machines on worker nodes, we implemented a prototype where each physical node provides several logical nodes (virtual machines) with different operating systems and various middleware and software configurations which can be used by users for running their jobs. Users may specify, using local submission system such as the PBS, which operating system configuration they would like to use. Some virtual machines can be used for submitting high-priority jobs which can suspend already running processes and steal their CPU and memory resources. All VMs running on one physical machine are coordinated by a special daemon running in the Domain-0 on each Xen-powered machine. The daemon maintains states of all virtual machines and assigns physical CPU and memory resources.

35. An Extensible Framework for Distributing Work in a Heterogeneous Environment using WebCom

Brian C. Clayton, Therese Enright, David A. Power, James J. Kennedy and John P. Morrison
Centre for Unified Computing, Dept. Computer Science, University College Cork, Ireland

WebCom provides a framework that supports distribution and execution of computational work in a distributed heterogeneous environment. WebCom is architecture neutral, modular in design and allows interaction with many computational platforms. Written in Java, WebCom is based on the Condensed Graph model of computing and provides a fault tolerant, secure, architecture neutral platform. Applications are expressed as Condensed Graphs. These consist of nodes and arcs, where nodes may be atomic instructions or condensed nodes potentially consisting of complete applications, such as MPI programs. Different types of arcs can be used to give different orders of evaluation: speculative, imperative and coercion driven. Workflows containing a mixture of these three orders of evaluation can be created easily and executed in heterogeneous environments. To date, WebCom has been deployed in Business To Business, Grid(WebCom-G) and MPI environments. Deployment in each new environment has necessitated the development of modules specific to that environment. Typically, such modules were not re-usable in different environments or user configurable, they were specific to the environment they were designed for.

The work presented here describes extensions to the WebCom framework based on the provision of user-defined policies, algorithms and meta-data in nodes. These extensions provide architecture-neutral mechanisms capable of handling user-defined job types that can potentially allow users to make optimum use of their resources and exploit resources currently unavailable to them. They are designed to be user configurable to ensure that work is allocated to resources according to policies and algorithms appropriate to the underlying platform.

The framework extensions described comprise a new Distributor Module which obeys distribution policies. These policies specify algorithms and distribution rules which can refer to both nodes and meta-data statically or dynamically attached to nodes. This meta-data is used to indicate the job type, for example a PBS job. Nodes passed to the distributor module are executed or distributed according to the specified policy.

A PBSDistributor module is currently being developed and will be deployed in a Torque/PBS managed environment as a proof of concept. This module can be employed to evaluate potential enhancements such as the ability of WebCom to back-fill unallocated nodes in a PBS managed environment, improving the overall utilisation of PBS managed clusters. WebCom could fill these underutilized nodes with non-PBS jobs such as Grid jobs, thereby informally opening up a cluster environment to Grid users without installing Grid software on the cluster.

The user-configurable environment within the distributor module ensures that Torque/PBS managed systems can be configured to accept work according to policies and algorithms considered optimal for the available resources. Extending WebCom to support multiple cluster resource management systems illustrates the potential of WebCom to operate as a meta-scheduler in a cluster, Grid, or `a cluster of clusters' environment.

36. Job Behaviour Characterization in Production-Level Computational Grids

D. Groen (1), A. Tirado-Ramos (1), D. Groep (2), W. van Leeuwen (2), J. Templon (2)
(1) Faculty of Sciences, Section Computational Science, University of Amsterdam, The Netherlands
(2) NIKHEF, Amsterdam, The Netherlands

Grids are often claimed to offer an infrastructure where there is always a site offering available resources for a job. This may be the case in theory, but the actual jobs that are submitted to the Grid do not always complete successfully. A variety of complications, obstacles and accumulated inaccuracies prevent users from running jobs in a significant number of cases [1]. It is widely reported in the literature that Grid usage offers more than 90% success rate for job submissions, though this is usually based on internal or unpublished statistics. Users, however, tend to experience a lower rate of success: some jobs never make it past submission, while others are successfully submitted but unsuccessfully executed [2]. This can cause scientists to submit an excess number of jobs to compensate for the high failure rate, leading to excess resource consumption. Previous efforts have been made to improve the stability of Grid network resources [3]. This work focuses primarily on the issue of Grid computational resource behavior characterization, from a user point of view. It aims to investigate the issues related to what factors can make a difference between job success and job failure. In contrast with previous theoretical approaches [4], we proceed by thoroughly examining a large quantity of real-world jobs with variable complexity and resource requirements. In doing so, we aim to explore the existing job success rate and the range of responses a particular computational site can exhibit. By tracking down the cause of failure for large numbers of jobs, we also distinguish common failure modes from less common ones.
We attempt to characterize the range of behavior of computational Grid sites. In order to do this in an accurate and broad fashion we created two classes of jobs (simple and complex) and attempted to submit them and execute the corresponding applications on a wide range of sites on different Grid infrastructures. In addition to that, we have zoomed in to a single site and tried to explore what information can be obtained at the system level. Before proceeding to the experimental setup and job analysis results, we have provided a very brief description of the infrastructures involved (LCG[5], CrossGrid[6]) and the kind of applications [7] we used for a given experiment. That is, a set of simple jobs that were submitted and analyzed after submission contained only a few trivial, low-level commands. Meanwhile, a series of more complex job analysis experiments were conducted, for which we closely monitored submissions. Furthermore, aside from user-perspective reliability measurements, we also took a closer look at specific Grid sites in order to determine how much information about job submissions can be retrieved from a local CE. For our experiments, we investigate and discuss the characterization of job behavior in production-level Grid environments. Our results show that site reliability is quite not a black-and-white issue. There have been numerous cases where sites cease functioning after behaving reliably, or vice versa. Also, it is not uncommon for sites to only respond reliably to some of the submitted jobs. Yet it also became apparent that site reliability is far from random. A number of sites behaved relatively reliably throughout our experiments, while others consistently showed a medium to high rate of failure. In addition, we have discovered a number of sites that were labeled as suitable in information systems, while delivering sub-optimal reliability in practice. These results indicate that it is useful to track general trends of site reliability in order to predict reliability for future job submissions.

Acknowledgements. This work was carried out in the context of the Virtual Laboratory for e-Science project (www.vl-e.nl), supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science (OC&W) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ).

References.
[1] Thain, D. Livny, M., "Error Scope on a computational grid, theory and practice", 11th IEEE International Symposium on High Performance Distributed Computing (HPDC-11), 2002, pp. 199-208.
[2] David Groep, Jeff Templon and Charles Loomis, "Crunching Real Data on the Grid: Practice and Experience with the European DataGrid", Concurrency and Computation: Practice & Experience, Volume 18,

Issue 9, 2006, pp. 925-940.
[3] Curti, C., Ferrari, T., Gommans, L., Van Oudenaarde, S., Ronchieri, E., Giacomini, F., Vistoli, C., "On advance reservation of heterogeneous network paths", Future Generation Computer Systems, Volume 21, Issue 4, 2005, pp. 525-538.
[4] Low C., "Decentralised application placement", Future Generation Computer Systems, Volume 21, Issue 2, 2005, pp. 281-290.
[5] The LHC Computing Grid, http://lcg.web.cern.ch [6] J. Gomes et. al., "Organization of the International Testbed of the CrossGrid Project", Cracow Grid Workshop 2005, Cracow, Poland, 2005, pp. 21-28.
[6] J. Gomes et. al., "Organization of the International Testbed of the CrossGrid Project", Cracow Grid Workshop 2005, Cracow, Poland, 2005, pp. 21-28.
[7] The D0 Experiment at Fermilab, http://www-d0.fnal.gov/

37. Preparing Storage Infrastructure to Meet the Requirements of the Grid Environment

M. Pogoda, G. Sulkowski, M. Twardy
Academic Computer Centre CYFRONET AGH, Kraków, Poland

Current evolving of the Grid Environment causes growing demand for storage resources implied by increasing number of data-intensive applications like those coming from HEP. Appropriate design of the storage infrastructure and its proper configuration may become the key element determining the performance of grid computing centers. In heterogeneous environments problems may also arise in the aspects of compatibility, scalability and flexibility. Grid storage service must balance widely divergent requirements: from high-throughput, high-available storage for strategic tasks, to high-capacity, low-cost disks for backups and infrequently used users data.

This article presents our experiences in resolving various problems encountered during the process of designing, implementing and optimizing the grid storage infrastructure at ACC CYFRONET AGH. Our activity has been aimed to explore proper way to improve performance storage disk array for heterogeneous environments. Effort has been also focused to provide methods for easier process management over storage infrastructure and to eliminate aspects of the time-consuming implementation process. Additionally we focused on some aspects which are normally not considered during installation process in typical storage infrastructure like:

- influence of the cache memory configuration for elements of the storage infrastructure in case of the central virtual configuration and central system management,
- differences in performance to date access between virtual and non virtual storage configuration
- dependency between communication processes in virtual configuration of the storage arrays.

The results of our investigation can be useful for preparing storage infrastructure to meet the requirements of the Grid Environment.

38. Monitoring and Analysis of Dependability Metrics of Grid Services

Robert Samborski, Hong-Linh Truong, Thomas Fahringer
Distributed and Parallel Systems Group, Institute of Computer Science, Innsbruck, Austria

The Grid gives a user access to a vast number of services, but it is often not easy to select the best fitting from many available, complement services distributed among many Grid sites. The dependability of once selected services can also vary greatly over time, making the service selection process challenging. Grid services must be constantly monitored to determine if they are dependable, that is if a user or middleware can entrust work to them.

This paper presents an architecture for monitoring, analyzing, visualizing and reasoning about dependability metrics of services that are available in a Grid. The architecture is flexible and scalable through utilization of a WSRF-based P2P middleware [1]. In our work we focus on monitoring a broad range of important, orthogonal dependability metrics characterizing typical Grid services. We develop sensors that monitor dependability metrics such as: availability of machines, Web Services, network paths, reliability of GRAM middleware or manageability of WSDM [2] compliant services.

Using rules and reasoning techniques, Grid services, their status and inherent casual relationships among these services can be automatically deducted from data provided by the middleware, and viewed in a dynamically generated graph. The user can define or extend monitoring scenarios by adding virtual intermediate services, modelled as group nodes, whose status is not directly monitored but computed depending on a set of other services, and by defining a set of additional casual and mutually exclusive relationships between existing services in the graph. This allows us to detect deficiencies in operational functionalities of available services and to extract additional data, e.g. reachability, from basic data, e.g.

availability. Our tool also supports different views at the monitored services, with each view having a different way of propagating service status in the graph.

In this paper we show a working prototype of the above mentioned architecture and illustrate the usefulness of the prototype by demonstrating real examples conducted in the Austrian Grid and the K-WfGrid testbeds.

Acknowledgments. The work described in this paper is partially supported by the European Union through the IST-2002-511385 project K-WfGrid and by the Austrian Science Fund through the SFBF1104 project Aurora.

References
[1] Hong-Linh Truong and Thomas Fahringer: Self-Managing Sensor-based Middleware for Performance Monitoring and Data Integration in Grids, In Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS2005), Denver, Colorado, USA, April 4-8, 2005.
[2] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsdm

39. Enhancing the Functionality of OCM-G and G-PM

Marian Bubak(1,2), Bartosz Balis(1), Wlodzimierz Funika(1), Marcin Radecki(2), Tomasz Szepieniec(2), Roland Wismueller(3), Pawe³ Marchewka(4), Jakub Janczak(1), Rafal Chojnacki(1)
(1) Institute of Computer Science AGH, Krakow, Poland
(2) ACC CYFRONET AGH, Krakow, Poland
(3) Universitaet Siegen, Germany
(4) Institute of Nuclear Physics PAN, Kraków, Poland

The successful development and usage of performance monitoring tools in the CrossGrid project, first of all the OCM-G monitoring system and the G-PM performance measurement tool motivates their further enhancements and use in adjacent with new programming paradigms and applications. These two facilities are originally oriented towards interactive grid applications which exploit the message passing paradigm, e.g. MPI. Monitoring of interactive applications is only possible in the on-line mode in which the information is immediately delivered to the visualization tools with low latencies. The monitoring infrastructure created in the CrossGrid project is the OCM-G which is a distributed, decentralized, autonomous system, running as a permanent Grid service and providing monitoring services accessible via a standardized interface OMIS, e.g. for the G-PM performance evaluation tool. The general idea of the tools is to provide/use a flexible set of monitoring services which allow to define metrics with a semantics relevant to the user needs with help of a metrics specification language, specifically PMSL, instead of providing a predefined, fixed set of high-level metrics.

In this paper, we provide an overview of the main trends of the evolution of the these tools. These enhancements are connected with new requirements put by the tools' use for the purposes different from those stemming from CrossGrid's specifics. Within the EU IST CoreGRID project, the OCM-G/G-PM are targeted to provide monitoring support for applications following the GRID superscalar paradigm. One of the interesting challenges in this case is to capture information on a vast amount of worker jobs which are spawned and completed once their goal is achieved. This poses the requirement to catch all changes in the application execution, especially to dynamically update the list of constituent processes. At the same time, performance metrics and visualization modes provided by the G-PM tool evolve towards developers' and end-users' needs, to a greater extent oriented to user-defined semantics. Within another project, the EU IST int.eu.grid, there is a requirement that a VO intended to support a pilot HEP application needs to get information on the status of tasks processing raw experimental data on distributed clusters, to make scheduling decisions. This kind of use requires access to low-level monitoring data as well as advanced instrumentation techniques. Other directions of development are realized in the frame of BalticGrid project [4], where OCM-G and G-PM is used to facilitate process of adapting scientific applications to grid environment. OCM-G itself was adapted to work in gLite middleware including simplified, automatic installation process. Answering users requirement the set of services was added to enable detailed supervising of remote nodes where user processes are running including spying output files on demand and spawning additional command on node controlled through OCM-G. Additionally, we made our tools more user friendly by simplifying instrumentation of MPI library and other features. Finally, within Clusterix project [5] OCM G and G-PM are adapted to monitor applications running on IA64 architecture within Globus 3.2 environment.

All initiatives listed above makes the OCM-G and G-PM one of the most mature, extended system for grid application monitoring.

Acknowledgements. This reseach is funded by the EU IST Projects CoreGRID, int.eu.grid, BalticGrid, and Polish Project CLUSTERIX as well as SPUB-M grants.

References

1. Balis, B., Bubak, M., Funika, W., Wismueller, R., Radecki, M., Szepieniec, T., Arodz, T., Kurdziel, M., Grid Environment for On-line Application Monitoring and Performance Analysis, Scientific Programming, vol. 12, no. 4, 2004, pp. 239-251.
2. Badia R.M., Bubak, M., Funika, W.and Smetek, M., Performance Monitoring of GRID superscalar applications with OCM-G, in: Sergei Gorlatch, Marco Danelutto (Eds.), Proceedings of the CoreGRID Workshop "Integrated Research in Grid Computing", November 28-30, 2005, Technical Report TR-05-22, 2005, pp. 229-236.
3. EU IST int.eu.grid project's web page: http://www.interactive-grid.eu/
4. Web page about OCM-G on EU IST BalticGrid project site:
http://www.balticgrid.org/Application_Support/Documents/ocmg.html
5. Clusterix Project web page: http://clusterix.pcz.pl/indexEng.php

40. Towards User-Defined Performance Monitoring of Distributed Java Applications

Wlodzimierz Funika, Piotr Godowski, Piotr Pegiel
Institute of Computer Science AGH, Krakow, Poland

The design of distributed application is in many cases a challenge to the developer. On the one hand, there are the limitations and performance issues of distributed programming platforms. On the other hand, the developer must assure that the application manages and uses distributed resources efficiently. Therefore, understanding application's behavior through performance analysis and visualization is crucial. Performance visualization tools are intended to support exploration of vast amounts of monitoring data on the distributed application with drilling down the available data and switching focus between performance aspects.

One of the most important tasks for the tool under discussion is to enhance performance visualization capabilities. Of course, there are numerous well developed tools that are used for monitoring applications. These tools frequently have quite a closed structure and it is rather hard to add new functionality in the future. Moreover, it often happens that a tool designer cannot foresee in what context this application will be used. Therefore it is very important to enable the user to add in a simple way own functionality, e.g. new metrics.

In this paper we focus on a visualization tool which provides an interface between raw data obtained from the monitoring system and various user's needs in performance analysis, based on the functionality of a Java oriented monitoring infrastructure - J-OCM. We use the on-line performance visualization approach - it helps understand how the application is working, how big is resources usage, and how different kinds of measured data are correlated with each other (e.g. synchronous communication overhead and total time of method execution). On-line visualization allows also for proper reactions e.g. if applications are requesting more resources. Additionally, due to using the J-OCM system it is possible to observe different applications within a single tool run. So, for example, we can track the behavior of a client and server and compare them with the same time line.

In the paper we present the design of Java Measurement Tool (JMT) intended to work with J-OCM monitoring system, and demonstrate its features, like types of performance visualization and user-defined metrics written in Java, which enables to define measurements and visualization meaningful in the context of application specifics. The intended goal is to support PMSL-based metrics and maximize the ease of: creating new metrics, aggregating metrics, user friendly visualization interface, and finally providing a set of basic metrics to support PMSL-compliant metrics.

References:
1. N. Podhorszki and P. Kacsuk. Presentation and Analysis of Grid Performance Data. In: Harald Kosch, László Böszörményi, Hermann Hellwagner (eds.), Parallel Processing, 9th International Euro-Par Conference, Klagenfurt, Austria, August 26-29, 2003. Proceedings. Lecture Notes in Computer Science 2790 Springer 2003, pp. 119-126
2. Daniel A. Reed and Randy L. Ribler. Performance Analysis and Visualization. In: Ian Foster and Carl Kesselman (eds), Computational Grids: State of the Art and Future Directions in High-Performance Distributed Computing, Morgan-Kaufman Publishers, August 1998, pp. 367-393.
3. W. Funika, M. Bubak, M.Smetek, and R. Wismuller. An OMIS-based Approach to Monitoring Distributed Java Applications. In: Yuen Chung Kwong, editor, Annual Review of Scalable Computing, volume 6, chapter 1. pp. 1-29, World Scientific Publishing Co. and Singapore University Press, 2004.

41. Secure Logistical Networking

L. Hejtmanek, L. Matyska, M. Prochazka
Masaryk University, Faculty of Informatics and Institute of Computer Science, Brno, Czech Republic

Logistical Networking is a distributed data storage system based on an Internet Backplane Protocol. It offers a universal storage fabrics that could be used in contemporary Grids to provide data storage, retrieval, and sharing services. However, legacy Logistical Networking offers no or very limited support for authentication and authorization, limiting severely it usability in the Grid environment. In this paper, we propose an architecture that extends Logistical Networking to use Grid authentication and authorization services.

The Logistical Networking system consists of three basic services---the metadata manager, that stores all the metadata related to individual files, the L-Bone server that keeps list of available storage servers, and the actual IBP servers. The metadata manager is our extension over the legacy Logistical
Networking systems where metadata were kept by clients. Our architecture guarantees that user is authenticated to all these services, the authorization granularity is also at the service level and all authorizations can be revoked at any moment by service providers (usually the storage owners). We also support access policies. These can limit maximum amount of distributed storage space allocated to a user or group of users or they can limit the maximum amount of time the client can keep his data within the distributed storage. Advanced access control to files is supported, administrators can define access conditions (e.g. time limit when a particular file can be accessed or a system load limit, when client can access his data only if server load is low). Also, file owner can bind these usage constraints with a file.

The authentication is based on X.509 certificates, the authorization uses VOMS attributes. We use user ID and group ID attributes to uniquely identify each user regardless of the number and nature of certificates she uses. Unlimited number of her certificates may be registered with the same user ID. The user ID mapping is part of the VOMS attributes that define authorization capabilities. To access the file, a time limited proxy certificate with VOMS attributes is used to check the authorization. The metadata and the list of storage servers are signed by the metadata managers and the L-Bone server, resp., and have a limited life time.

During all the authentication and authorization step, the communication is always between client and one service only, this removes a risk of an access refusal due to some hidden remote service unavailability. Also, no revocation lists are necessary, as all the rights are time limited and therefore not cache-able.

The prototype implementation has been used to evaluate overhead associated with the security enhancement. This overhead is negligible for Auth, Allocate, Load, and Store commands if only capabilities are encrypted, the Copy command has a penalty of some 10ms (rather negligible for large data transfers). When full data encryption is enforced, the Load, Store, and Copy command are bound by the speed of the used AES 128 cipher.

42. Specification, Analysis and Testing of Statefull Grid Services Using Abstract State Machines

Daniel Lamch, Roman Wyrzykowski
Institute of Computer and Information Sciences, Czestochowa University of Technology

Abstract State Machines (ASM) are mathematically defined environment for high-level system design, verification and analysis. This paper presents a proposition of a hybrid approach to the specification, analysis and testing of statefull grid services based on WSRF (Web Service Resource Framework) using ASM. This approach allows for easy integration of specification of developed middleware with existing elements of grid systems.

The important advantage of the approach is an automatic generation of test procedures for the implementation, following the model-based testing approach. It allows for a smooth transition from the specification to implementation stage, as well as for examination of features of specification and implementation at every stage of their development.

Also, we propose a software environment which implements he proposed approach. Its use in practise will help to create more reliable grid systems.

Keywords:
grid systems, middleware, specification, testing, WSRF, Abstract State Machines

43. TeToN - a Jena-based Tool for text-to-ontology Approach

Marcin Kuta (1), Stanislaw Polak (1), Barbara Palacz (1) , Tomasz Milos (1), Renata Slota (1), Jacek Kitowski (1,2)
(1) Institute of Computer Science, AGH, Krakow, Poland
(2) Academic Computer Center CYFRONET AGH, Krakow, Poland

Current development of grid environments and grid computing are focused on extensive use of knowledge for efficient exploitation of both hardware and software grid resources, with ontological formalism for knowledge representation applied typically. Good examples in this field are EU IST projects, like OntoGrid [1], Inteligrid [2] and K-WfGrid [3]. In the last mentioned project dynamically gathered knowledge is used for evaluation the application workflows that are composed dynamically with WSRF services selected automatically according to a user request. One way of expressing evaluation score given to the user is to use notes in natural language, which, later on can be presented as a tip to another one being in the same context of work.

In this paper we present a spin-off of the research performed within K-WfGrid, which consists of making knowledge of the tips more useful formally. For this purpose we propose to study creation and extension of existing ontologies by understanding natural language sentences. Since input sentences provide additional feedback from the user, it may be helpful for improvement of grid operation. The previous studies on natural language text to ontology approach focus on methods that require large text corpora [4] and extract dependencies by means of regular expressions [5].

The TeToN (Text To oNtology) tool is intended to acquire new knowledge by analysing only short pieces of information in form of sentences in English, in the future devoted to the problems addressed by projects like K-WfGrid, making use of their existing system knowledge.

Understanding language means to know what concepts a word or phrase stands for and to know how to link those concepts together in a meaningful way. In order to perform this task the input sentence is converted to ontology elements. Several steps are performed: syntactic and semantic analysis, ontology binding, relation matching and ontology update. During the first two steps potential concepts, relations and instances are identified. These elements are confronted with ontology, for which an index is created. The index associates list of WordNet keywords with concepts from ontology.

For each concept a measure indicating the accuracy of the match to ontology is calculated. Only concepts with measure above certain threshold are further considered. Next concepts are bind to proper relations. Having selected concepts and matching relations the ontology is extended. Optionally new relation entry is created if it does not exist in dictionary.

The Jena library is used to manipulate the knowledge model based on an OWL. Other subtasks like word sense desambiguation or ontology indexing also required dedicated tools and libraries.

[1] OntoGrid website: http://www.ontogrid.net/
[2] Inteligrid website: http://www.inteligrid.com/
[3] K-Wf Grid website: http://www.kwfgrid.eu/
[4] Liu, W., Weichselbraun, A.,"Semi-Automatic Ontology Extension Using Spreading Activation", Journal of Universal Knowledge Management, vol. 0, no. 1 (2005), 50-58.
[5] Hearst, M.A., "Automatic acquisition of hyponyms from large text corpora", In Proceedings of the 14th International Conference on Computational Linguistics, pages 539--545, 1992.

44. Advanced Mechanisms for Metadata Storage and Synchronization in CDMS^2

Lukasz Kuczynski, Konrad Karczewski and Roman Wyrzykowski
Institute for Computational and Information Sciences, Czestochowa University of Technology, Poland

Grid applications deal with large volumes of data, which should be accessed in a fast and secure manner. Consequently, effective data management solutions are vital for the success of grid technology. The CLUSTERIX Data Management System (CDMS) has been developed in the CLUSTERIX Polish grid project, based on the analysis of both the end-user's requirements and existing implementations. During its development, a special attention has been paid to making the system reliable and secure, aiming at the same time at creation a user-friendly grid data storage system.

In the system development phase, many concepts were proposed and postponed due to a limited time-frame of the project. All those improvements were intended to be implemented during future work on the CDMS project, but after careful consideration it was decided that it would be easier to re-implement the entire system taking into account experience gathered during CDMS implementation. This decision led to the creation of $CDMS^2$.

The main difference between the existing system and $CDMS^2$ would be metadata storage method. Currently a relational database is used but this solution led to unexpected problems with data coherency in an environment with distributed Data Broker. To avoid them in the new implementation a dedicated filesystem was proposed: cdmsfs.

The architecture of this filesystem would be designed for operation in a distributed environment by including data replication mechanism and optimized for $CDMS^2$ metadata storage. The cdmsfs will include a statistics gathering mechanism which would allow to eliminate the statistical subsystem from the first version of CDMS. This will contribute to the creation of a consistent mechanism ensuring high system reliability.

This paper presents the advanced metadata storage and synchronisation mechanisms implemented in the data management system CDMS^2.

CLUSTERIX Project Home Page, http://clusterix.pcz.pl

45. Building gLite Based Scalable Grid Environment with HP SFS

Peter Dobe, Denes Nemeth, Imre Szeberenyi
Budapest University of Technology and Economics, Hungary

In most cases it is very difficult to maintain and administer large number of computers formed grid site.

We have created an NFS or/and SFS based solution, which allows nodes to be added in a matter of minutes without prior installation of gLite software. The worker nodes are nearly diskless: most part of the file system on each is served via NFS root located on the GG host. This supplies all the necessary applications and configuration files to operate in the Grid environment. Only temporary data files and the host-specific private keys and certificates are stored locally. The latter two are required by the PKI based authentication mechanism that is commonly used in Grids such as the EGEE Grid.
It is also possible to deploy completely diskless nodes where these are downloaded from a secure network. The nodes system can served by NFS or SFS.

Our HP Proliant G2 worker nodes are linked with a Gigabit Ethernet network, which also connects them to an HP Scalable File Share (SFS) storage system of approximately 3 Terabytes capacity. The site is part of the EGEE infrastructure, and runs the gLite middleware.

The homogeneous set of the HP servers makes it possible to use one common kernel image on each host without further configuration on the hosts separately.

The SFS is based on the Lustre File System, and provides an efficient administration environment and a single point of management. The flexible Lustre technology permits a huge variety of configurations and allows grid administrators to fine tune the system to meet the specific needs of the different types of applications.

SFS provides a network-independent solution with high network performance, redundancy, higher availability and transaction rates than the standard solutions.

In addition to being part of the European Grid, the site serves as a computing cluster for the computations at BME Faculty of Architecture. The problem being solved is calculating the prestressing strength of reinforced concrete bars used in bridges.