A Platform for Collaborative e-Science Applications

Marian Bubak[1, 3], Tomasz Gubała [3, 2], Maciej Malawski[1], Bartosz Balis[1],
Włodzimierz Funika[1], Tomasz Bartyński[3], Eryk Ciepiela[3], Daniel Harężlak[3],
Marek Kasztelnik[3], Joanna Kocot[3], Dariusz Król[3], Piotr Nowakowski[3],
Michał Pelczar[3], Matthias Assel[4]

[1]Institute of Computer Science AGH, al. Mickiewicza 30,
30-059 Krakow, Poland
[2]Informatics Institute, University of Amsterdam, Kruislaan 403,
1098 SJ Amsterdam, The Netherlands
[3]ACC CYFRONET AGH, Krakow, ul. Nawojki 11, 30-950 Krakow, Poland
[4]HLRS High Performance Computing Center Stuttgart, University of Stuttgart, Nobelstrasse 19
70550 Stuttgart, Germany

Abstract

A novel, holistic, approach to scientific investigations should, besides analysis of individual phenomena, integrate different, interdisciplinary sources of knowledge about a complex system to obtain a deep understanding of the system as a whole. This innovative way of research, recently called system-level science [1], requires advanced software environments to support collaborating research groups. Most problem-solving environments and virtual laboratories [2] are built on top of workflow systems [3]. Their main drawbacks include limited expressiveness of the programming model and lack of mechanisms for integration of computing resources from grids, clusters, dedicated computers, and – recently – clouds.

In the ViroLab project [4] we are developing a virtual laboratory [5,6] to support research of infectious diseases, facilitate medical knowledge discovery and provide decision support for HIV drug resistance studies [7]. This virtual laboratory may also be used in other areas of system-level science [8]. Within ViroLab, Collaborative e-Science applications are considered experiments. The Experiment plan notation is based on the Ruby scripting language which has a concise and clear syntax combined with a full set of control structures, enabling relatively easy construction of experiments with various degree of complexity. To provide easy access to different computing and data resources, we have introduced a grid object abstraction level hierarchy. A grid object class is an abstract entity which defines the operations that can be invoked from the script. A class may have multiple implementations, representing the same functionality, and each implementation may have multiple instances, which may run on different resources. Grid objects may have different properties, such as stateless or stateful interaction modes, and they may be either private or shareable between experiments and users.

The Virtual Laboratory (see Fig 1) is equipped with tools for user-friendly experiment creation and execution, enabling reuse of existing experiments, gathering and exploiting provenance information, as well as integration of geographically-distributed compute and data resources.

The Experiment Planning Environment supports rapid experiment plan development while the Experiment Management Interface enables loading and execution

of experiments. The Experiment Repository stores experiment plans prepared by developers and published for future use. The virtual laboratory engine includes the Grid Operation Invoker which instantiates grid object representatives and handles remote operation invocations. The GridSpace Application Optimizer is responsible for optimal load balancing on computational servers. The Data Access Service acquires data from remote databases located in research institutions and hospitals.
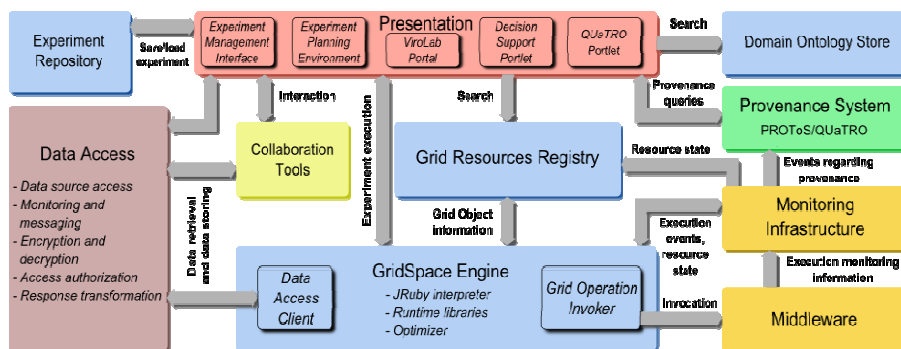
Fig. 1. Architecture of the Virtual Laboratory

The provenance approach in the ViroLab virtual laboratory brings together ontology-based semantic modeling, monitoring of applications and the runtime infrastructure, and database technologies, in order to collect rich information concerning the execution of experiments, represent it in a meaningful way, and store it in a repository.

In the ViroLab project, this virtual laboratory is used to plan and execute important virological experiments, with various types of analysis of the HIV virus genotype, such as the calculation of drug resistance, querying historical and provenance information about experiments, a drug resistance system based on the Retrogram ruleset. It has also been applied to other application domains, such as: protein folding and structural comparison, data mining using the Weka library, computational chemistry – to develop a series of Gaussian application on the EGEE infrastructure. as an education tool in computer science classes.

We have developed an environment for collaborative planning, development and execution of e-Science applications. It facilitates fast, close cooperation of developers and users so it may be used by groups of experts running complex computer simulations. In-silico experiments undergo frequent changes, modifications and enhancements, and this platform encourages quick, agile simulation software releasing.

Acknowledgments

References

1. I. Foster and C. Kesselman: "Scaling System-Level Science: Scientific Exploration and IT Implications", *IEEE Computer*, vol. 39, no 11, 31-39, 2006.
2. K. Rycerz, M. Bubak, P.M.A. Sloot, V. Getov: "Problem Solving Environment for Distributed Interactive Simulations" in: Sergiei Gorlatch, Marian Bubak, and Thierry Priol (Eds): *Achievements in European Reseach on Grid Systems.* CoreGRID Integration Workshop 2006 (Selected Papers) ISBN-13: 978-0-387-72811-7; pp 55 - 66, Springer, 2008.
3. Y. Gil, E. Deelman, M. Ellisman, T. Fahringer, G. Fox, D. Gannon, C. Goble, M. Livny, L. Moreau, and J. Myers; "Examining the Challenges of Scientific Workflows". *IEEE Computer* vol. 40, no 12, pp 24-32, 2007.
4. ViroLab – EU IST STREP Project 027446; `www.virolab.org`.
5. The ViroLab Virtual Laboratory, `http://virolab.cyfronet.pl`.
6. M. Bubak, T. Gubala, M. Malawski, B. Balis, W. Funika, T. Bartynski, E. Ciepiela, D. Harezlak, M. Kasztelnik, J. Kocot, D. Krol, P. Nowakowski, M. Pelczar, J. Wach, M. Assel, A. Tirado-Ramos: „Virtual Laboratory for Development and Execution of Biomedical Collaborative Applications", S. Puuronen, M. Pechenizkiy, A. Tsymbal, D-J. Lee (eds.), Proc. 21st IEEE International Symposium on Computer-Based Medical Systems, June 17-19, 2008, Jyvaskyla, Finland, p. 373 - 378, DOI 10.1109/CBMS.2008.47.
7. P. M.A. Sloot, I. Altintas, M. Bubak, Ch. A. Boucher: "From Molecule to Man: Decision Support in Individualized E-Health", *IEEE Computer* vol. 39, no 11, 40-46, 2006.
8. T. Gubala, M. Kasztelnik, M. Malawski, M. Bubak: "Towards System-Level Science Support", in: M. Bubak, G. D. van Albada and J. Dongarra and P. M.A. Sloot (Eds.), *Proceedings of Computational Science - ICCS 2008*, 8th International Conference Krakow, Poland, June 2008, volume I, LNCS 5101, Springer, 2008, pp. 56-65.