The PL-Grid Virtual Laboratory in the Life Sciences Domain

Maciej Malawski, Eryk Ciepiela, Tomasz Gubała, Piotr Nowakowski, Daniel Harężlak, Marek Kasztelnik, Joanna Kocot, Tomasz Bartyński i Marian Bubak
ACC Cyfronet AGH, ul. Nawojki 11, 30-950 Kraków, Poland
e-mail: e.ciepiela@cyfronet.pl

Modern life sciences, particularly simulations in biochemistry, genetics and virology, impose significant requirements on underlying IT infrastructures. Such requirements can be loosely grouped into two general domains: demand for computational resources and demand for new software tools facilitating effective, productive and collaborative exploitation of such resources by a vast range of beneficiaries. While the key goal of supporting scientific experimentation with computerized infrastructures remains the provision of large-scale computational and data storage facilities, it is equally important to supply scientists with tools which enable them to collaboratively develop, share, execute, publish and reuse virtual experiments.

The presented Virtual Laboratory, currently under development within the scope of the PL-Grid project aims to respond to these requirements by supplying software which permits the execution of virtual experiments written in popular scripting languages on the distributed resources provided by HPC institutions participating in the project. The goal of the Virtual Laboratory is to a propose a model and facilities for exploratory, incremental scripting – already omnipresent in e-scientific research – and make it reusable and actionable for entire communities. It aims to bridge the gap between the oft-inaccessible high performance computing infrastructures and the end user's (i.e. domain scientist's) desktop computer, typically used to run calculations and collate experimental data. Rather than persuade the scientists to change their daily habits, we wish to provide an environment which meshes seamlessly with their style of work, yet extends their experimentation and collaboration potential with the capabilities of high-performance computing clusters.

Our experience gathered through development of the ViroLab Virtual Laboratory for virologists [VL, VL-ROYAL, VL-WWW], the APPEA runtime environment for banking and media applications in GREDIA project [APPEA, GREDIA-WWW] as well as the GridSpace environment [GS-WWW] for running in-silico experiments, has been augmented with user requirement analysis conducted during the initial phase of PL-Grid project, involving groups of scientists from various domains such as physics, chemistry and biology.

Within the PL-Grid Virtual Laboratory individual applications are referred to as experiments, which means that they go beyond simple and repeatable execution of installed programs. Indeed, they may involve experimentation or exploratory programming, where the application consists of multiple steps which are not known in advance and often are selected ad-hoc, based on the results of previous steps.

In order to meet these objectives the Virtual Laboratory follows a layered and distributed architecture, which begins with a web Portal (an entry point for the whole Virtual Laboratory) where users can access the Virtual Laboratory from any workstation equipped with a web browser. The portal, in turn, exposes the Experiment Workbench Layer – a common tool-rich workbench for all Virtual Laboratory researchers where they perform their daily experimentation, collaborate, communicate and share resources (such as reusable code written in a number of popular scripting languages, including Python, Perl and Ruby). Further below lies the Experiment Execution Layer where individual parts of experiments (also called snippets) provided by the users through the Portal are evaluated in the context of a particular user account on the so-called experiment host machine. The Virtual Laboratory also encompasses a layer of gems, i.e. resources available to the PL-Grid community, including PBS clusters, Grids, external services (e.g. Web Services) and data sources (e.g. RDBMSs) that users may make use of in the course of their research activities.

PL-Grid provides a wide array of scientific software which can be exploited by its users through the Virtual Laboratory. The offer includes biological packages (AutoDock, BLAST, ClustalW2, CPMD, Gromacs, NAMD); quantum chemistry applications (ACES II, ADF, CFOUR, CHARMM, Dalton, GAMESS, Gaussian, MacroModel, Molcas, Molpro, MOPAC, NWChem, OpenBabel, Siesta, TURBOMOLE) as well as mathematics

and physics-oriented tools (Mepp, Mathematica, MATLAB). The project consortium is always on the lookout on potential new users and extends an invitation for the Polish scientific community to take advantage of the PL-Grid computing infrastructure and software resources.

Acknowledgements
The research presented in this paper has been partially supported by the European Union within the European Regional Development Fund program no. POIG.02.03.00-00-007/08-00 as part of the PL-Grid project (www.plgrid.pl) and ACC Cyfronet AGH grant 500-08.

References

[APPEA] D. Harezlak, P. Nowakowski, M. Bubak, "Appea: A Framework for Design and Implementation of Business Applications on the Grid", Proceedings of Computational Science - ICCS 2008, 8th International Conference Krakow, Poland, June 2008, volume III, LNCS 5103, Springer, 2008

[GREDIA-WWW] The GREDIA project, http://www.gredia.eu

[GS-WWW] The GridSpace portal, http://gs.cyfronet.pl

[VL] M. Bubak, M. Malawski, T. Gubala, M. Kasztelnik, P. Nowakowski, D. Harezlak, T. Bartynski, J. Kocot, E. Ciepiela, W. Funika, D. Krol, B. Balis, M. Assel, and A. Tirado Ramos. Virtual laboratory for collaborative applications. In M. Cannataro, editor, Handbook of Research on Computational GridTechnologies for Life Sciences, Biomedicine and Healthcare, chapter XXVII, pages 531-551. IGI Global, 2009.

[VL-Royal] P.M.A Sloot, Peter V. Coveney, G. Ertayalan, V. Mueller, C.A. Boucher, and M. Bubak: HIV decision Support: from Molecule to Man. Philosophical Transactions of the Royal Society A, vol 367, pp 2691 - 2703, 2009, doi:10.1098/rsta.2009.0043.

[VL-WWW] ViroLab Virtual Laboratory, http://virolab.cyfronet.pl