

InSilicoLab: From generic solutions to an application-specific science gateway

Tuesday, 9 April 2013 11:55 (15 minutes)

Impact

The InSilicoLab portals may serve as an everyday workspace for scientists from different domains of science. As a tool accessible through a Web interface, it also enables researchers from any place in the world, to collaborate on a common research task, represented by an in silico experiment; and to use available large computing power to realise this task. In this way, the InSilicoLab portal may become a science gateway for a researchers community.

Currently, three portals - different instances of the InSilicoLab environment - run for two different communities in Astrophysics and one in Computational Chemistry. The applications of the environment vary from Molecule Optimisation in chemistry, to telescope simulation or molecular-hydrodynamics simulations in Astrophysics. The tool was already validated and used by researchers from the the aforementioned domains.

For the chemists, it enabled performing, so called, conformation scans as well as simple access to chemistry software - Gaussian, GAMESS and TURBOMOLE installed on the grid infrastructure. The portal provides also a specialised application enabling preparation of a chemistry application input - extracting the relevant information from larger input files.

In the Astrophysics domain, the portal is used for performing Monte Carlo simulations with use of custom software developed especially for this purpose and installed on the grid. The use of InSilicoLab allows performing parameter study to decide on specific parameters of telescopes that will be built for the Cherenkov Telescope Array (CTA) observatory. Such a study is necessary for the design of individual telescopes included in the array.

The second Astrophysics portal is designed to perform simulations using a dedicated Molecular Hydro-Dynamics application.

In both aforementioned domains, stress was put on managing the data obtained in the course of experimentation, which tends to be very complex. The InSilicoLab environment provides a workspace that allows to organise all the researchers' data properly, as well as to use it for further processing. A built-in LFC catalogue manager enables also sharing large input files used by the community for their research. The computation processes can be tracked organised as well to enable even easier experiment management.

URL

<http://insilicolab.grid.cyfronet.pl>

Summary

A temptation of making the computer systems of today as powerful as possible led software architects to building very generic systems, capable of doing different operations on many computational platforms.

However, the authors observed that no matter how powerful such systems would be, they usually are unintuitive to their users, who, in reality, need to focus only on a small part of the tool capabilities that would help them to solve their scientific problems and keep track of their everyday work.

This observation was later confirmed by popularisation science gateways, which are tailored to a specific user community.

The InSilicoLab environment is a model which instances form gateways specialised for users from different domains of science. Their interface as well as processing routines are fully customised to work of scientists from a concrete domain or community. At the same time, the architecture of the system remains generic enough to make it powerful and its code maintainable.

Description

Motivation:

The scientists who use large computational power such as clusters, Grids or Clouds usually follow certain paths in the simulation they run everyday. They need a specific functionality that would assure their computations are defined properly, the data processed and all the relevant results stored for further use.

What they are offered, are usually large, complex systems that would allow them to define their computation in a format usually unknown and uncommon to the scientist; or a set of individual specialised tools that are difficult to be combined or run in different environments (e.g. on computing clusters).

What they need, is usually the ability to use their specialised applications, but on larger resources and in larger workflows that would enable them to go from the input data to the most relevant results, not being obstructed by the need for format conversion, data transfer or the technological knowledge necessary to run the computation.

Implementation:

The InSilicoLab environment offers a scientist a ready work environment that on one hand has an interface resembling the specialised tools familiar to the users, and on the other hand, conduct the whole experimentation process, using large computing resources and parallelisation where necessary.

The two layers - the specialised domain-specific tools and the computing resources are bridged by the middle-layer components transforming a domain-specific scenario of a scientific computation to an execution on the computational infrastructure. They are also responsible for tracking the provenance of the data produced or input into the experiment, assuring repeatability and retraceability of the researcher's computation and enabling composition of larger workflows that join many computations in a repeatable set of actions.

Primary authors: KOCOT, Joanna (CYFRONET); STERZEL, Mariusz (CYFRONET); SZEPIENIEC, Tomasz (CYFRONET)

Presenter: KOCOT, Joanna (CYFRONET)

Session Classification: Science Gateways Workshop

Track Classification: Virtual Research Environments (Track Lead: G Sipos and N Ferreira)