GridSpace Based Virtual Laboratory for PL-Grid Users

Maciej Malawski, Eryk Ciepiela, Tomasz Gubała, Piotr Nowakowski, Daniel Harężlak, Marek Kasztelnik, Joanna Kocot, Tomasz Bartyński and Marian Bubak
ACC CYFRONET AGH

Outline

- **Motivation** Shift from traditional to *e-science.* IT supporting science in *in-silico* research
- **Problem Definition** How to accompany scientists in their *actual daily workflow* and conceal complexity and low-level issues of underlying IT
- **Scientific and Technological Challenges** Notion, model and technology for creating, describing, (re) enacting, publishing and sharing actual daily workflows that involve IT
- **State of the Art** Virtual Laboratories. GridSpace-based ViroLab Virtual Laboratory
- **Description of the Solution** GridSpace2 – platform supporting complex workflow-like applications over computational infrastructure that involve experimentation and exploratory, dynamic and collaborative development.

Motivation

- Shift from traditional science to *e-science*
- Significant learning curve for scientist to master IT skills
- ...but IT is about leveraging abstraction level
- Software tools expected to
  - conceal complexity and low-level issues of underlying technology
  - accompany scientists with a workbench tailored to their demands and their actual style of work
- Experience and observations already gathered
  - GridSpace Virtual Laboratory for virologists in ViroLab project
  - APPEA runtime environment for banking and media application in GREDIA project
  - user inquiries, meetings during the user requirements analysis phase of PL-Grid project
  - collaboration with local research groups (virology, chemistry)

Problem Definition

- Scientist during their daily work deal with complex workflows that they go beyond simple and repeatable execution of installed programs
- Such workflows 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
- Workflows are conducted by ad-hoc research groups with collaborative participation of group members.
- Once enacted, workflows may need to be re-enacted again and again many more times. However, some indispensible adaptation and customization need to be made ad-hoc, dynamically - while the workflow enactment is already started.
- Workflows need to be conducted under scientist's continuous supervision and are subject to scientist's validation and even intrusion. Workflows cannot be fully automated.
- Workflows and their subparts, methods, libraries, ticks (!) involved are weakly described and belong to the individual scientist's know-how. They are not shared, nor published, not even well-specified.
- Workflows involve steps realized on various software platforms (not just written in many programming languages)

Scientific and Technological Challanges

- **Scientific** To develop notion/model for software engineering paradigm that is suitable for conducting scientific research that would embrace:
  - experimentation,
  - exploratory programming,
  - dynamic, ad-hoc programming,
  - Multiple software development platforms use
  - collaborative development,
  - reusability of whole experiments, their parts, libraries and services
  - re-enactment of experiment under executor's continuous supervision, ad-hoc intrusion, validation.
  - publishing of experiments as a service with specified terms of use.
- **Technological** To enable such paradigm with existing IT
  - computational infrastructures (clusters, grid, cloud)
  - software programming platforms actually used by target groups of scientists
  - infrastructure and administration of PL-Grid

State of the Art

- **GridSpace**
  - Already available technology developed by ACC Cyfronet
- **Virolab Virtual Laboratory**
  - Instance, deployment of GridSpace dedicated to ViroLab project
- **APPEA**
  - Instance, deployment of GridSpace dedicated to Gredia project
- **GridSpace 2**
  - Emerging technology (2.5 months of development preceded by analysis and design phases) developed by ACC Cyfronet
  - New incarnation of GridSpace 1 (new approach to the same problem)
  - Installable on arbitrary computational infrastructures
- **PL-Grid Virtual Laboratory**
  - Instance, deployment of GridSpace 2 technology dedicated to PL-Grid
  - Integrated with PL-Grid security, administration etc.

Description of the Solution – Vision

GridSpace 2

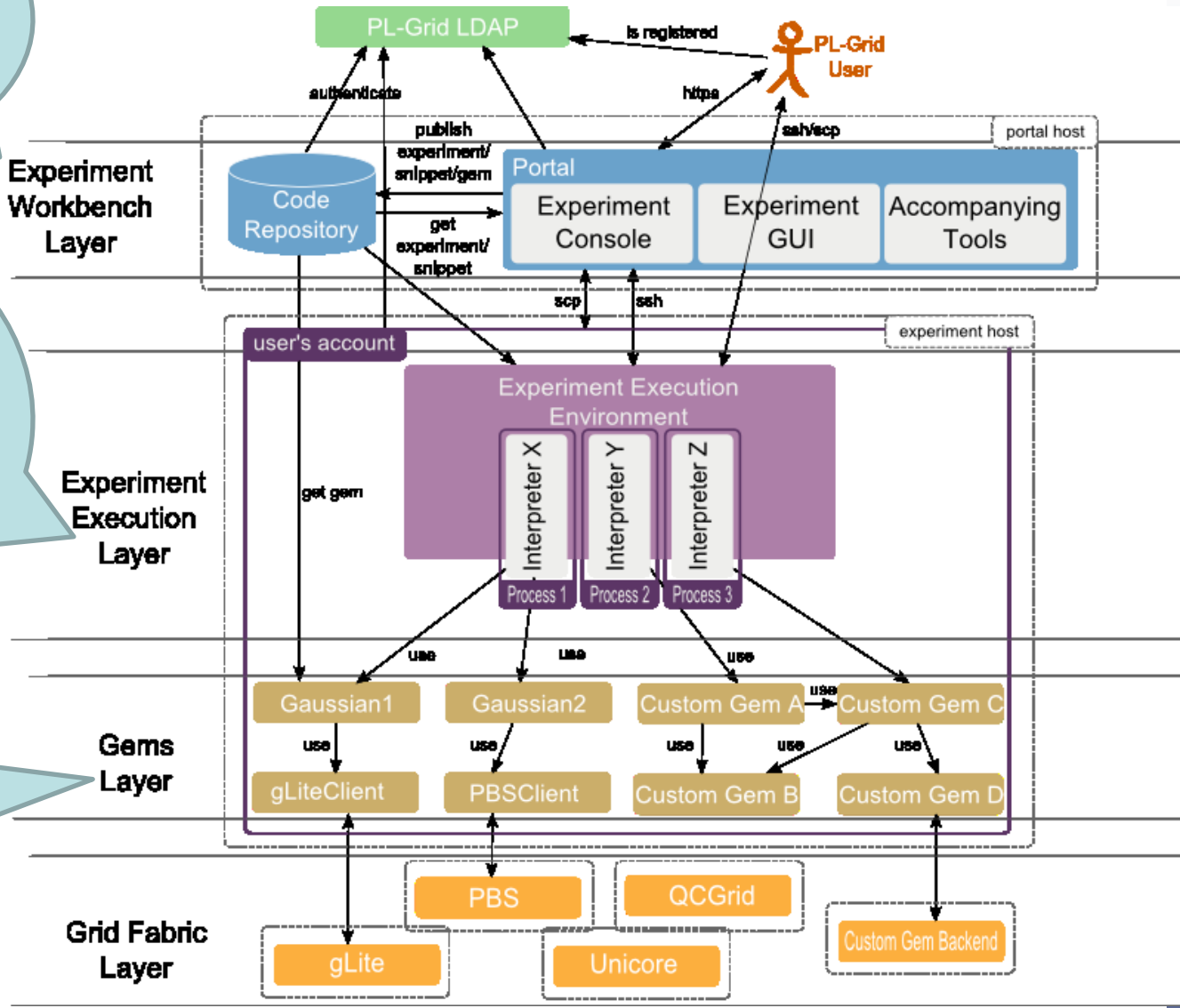
- Platform facilitating programming and execution of complex applications (*experiments*), that go beyond simple and repeatable execution of installed programs
- Experiments involve experimentation and exploratory programming
- Experiments consist of multiple steps often not known in advance and selected ad-hoc, based on the results of previous steps
- Experiments are subject to collaborative work – they can be developed, shared and reused amongst ad-hoc researching teams which collaboratively owns the libraries and services used by applications (*gems*), experiment parts (*snippets*) and whole experiments.

Description of the Solution

Portal, web tools, collaboration, reusing…

Experiment hosts (e.g. ui.grid.cyfronet.pl), interpreters (Python, Ruby, Bash, Perl…)

Libraries, utilities access to services



Current Achievements

- User's requirements analysis
  - Decided to carry out continuously during whole PL-Grid project
    - So talk to us and share your opinions
  - User inquires analyzed
  - User meetings carried out
- Design
  - Decided to allow for modification during whole PL-Grid project
  - First iteration (until mid 2010) designed completed
- GridSpace2 platform prototype – premiere now (PL-Grid account needed) – supports so far:
  - File management
  - Basic experimentation
  - PBS gem
  - Gaussian gem
  - Jmol viewer
- Your feedback needed to keep making it better

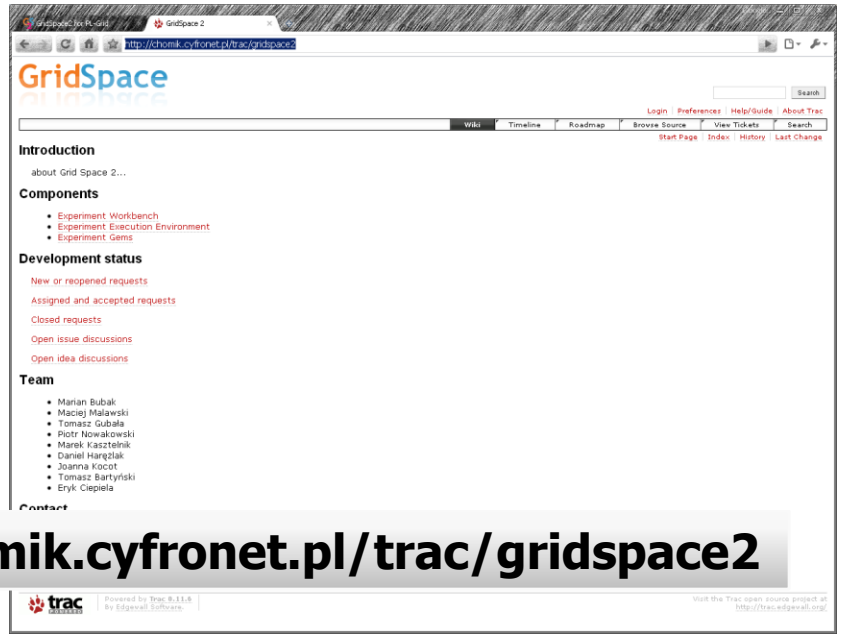
Future Work

- Now – First Prototype, subsequent versions will be continuously deployed on https://wl.plgrid.pl
- **Incremental development**
- IV.2010 – Second Prototype
- VI.2010 – First Release
- X.2010 – Second Release
  (to be presented at Cracow Grid Workshop 2010)
- XII.2010 – First fully integrated deployment on PL-Grid infrastructure
- 2011 - Further development

Conclusions

https://wl.plgrid.pl

For users…



…for feedback, contribution

https://chomik.cyfronet.pl/trac/gridspace2