Support for Collaborative e-Science Applications

Marian Bubak and Adam Belloum

Spiros Koulouzis, Regginal Cushing, Tomasz Gubala, Jakub Moscicki, Katarzyna Rycerz

Informatics Institute, University of Amsterdam, The Netherlands

www.science.uva.nl/~gvlam/wsvlam/

University of Amsterdam — Computational Science

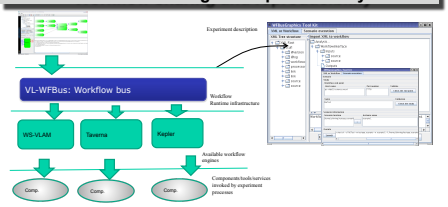
Motivation

- **Computing environments**
  - Distributed and heterogeneous
  - Resources shared between different organizations
  - Resources may dynamically change
  - No single middleware
  - Collaborations in virtual organizations
- **Scientific applications**
  - Collaborative
  - Used in dynamic scenarios
  - Compute- and data-intensive
  - Multiscale, multiphysics
  - Various levels of coupling and composition types
  - Legacy codes in many programming languages
  - Linking to publications

objectives

- Investigations aimed at development of problem solving environments to facilitate programming and execution of complex cross-disciplinary e-Science applications
- Providing support for collaborative aspects of research through dedicated tools for different groups of users: administrators, experiment developers, scientist
- Development of methods and tools for sharing and reuse of applications components
- Elaboration of a methodology of gathering and storing information about users' experience by tracing their behavior
- Understanding of mechanisms governing complex e-infrastructures for reliable execution of applications

Workflow sharing and reproducibility



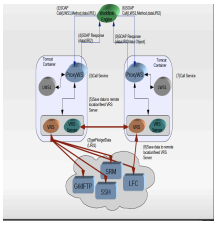
Zhiming Zhao, Adam Belloum, and Marian Bubak **Support for collaborative experiments in VL-e: from scientific workflows to knowledge sharing**, 4th IEEE international conference on e-Science and Grid Computing, Indianapolis, USA, 2008

Adam Belloum, Marcia A. Inda, Dmitry Vasunin, Vladimir Korkhov, Zhiming Zhao, Han Rauwerda, Timo M. Breit, Mariabak, Luis O. Hertzberger, "Collaborative e-Science Experiments and Scientific Workflows," IEEE Internet Computing, vol. 15, no. 4, pp. 39-47, July/Aug. 2011, doi:10.1109/MIC.2011.87S

Scientific data management

Scalable Data Transfers for Web Services

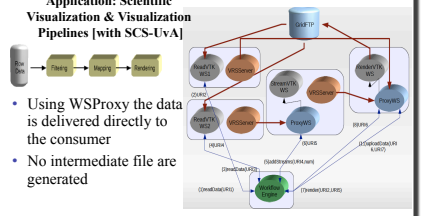
- Invocation are redirected to the ProxyWS which returns references instead of large data sets.
- Transparent access to divers data recourses



Spiros Koulouzis, Edgar Meij, M. Scott Marshall and Adam Belloum "**Enabling Data Transport between Web Services through alternative protocols and Streaming**" 4th IEEE international conference on e-Science and Grid Computing. Indianapolis, USA, 2008

Scientific data management

Application: Scientific Visualization & Visualization Pipelines [with SCS-UvA]



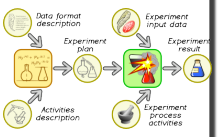
- Using WSProxy the data is delivered directly to the consumer
- No intermediate file are generated

Spiros Koulouzis, Elena Zudilova-Seinstra, Adam Belloum, "**Data Transport between Visualization Web Services for Medical Image Analysis**" *International Conference on Computational Science, ICCS 2010.*

Cooperative Virtual Laboratory for e-Science

Design of a laboratory for virologists, epidemiologists and clinicians investigating the HIV virus and treating HIV-positive patients

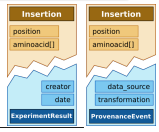
- Full concept-prototype-refinement-production circle for virology tools
- Set of dedicated yet interoperable tools bind together programmers and scientists for a single task
- Support for system-level science with concept of result reuse between different experiments

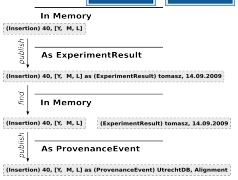


T. Gubala, M. Bubak, P.M.A. Sloot, "Semantic Integration of Collaborative Research Environments", chapter XXVI in "Handbook of Research on Computational Grid Technologies for Life Sciences, Biomedicine and Healthcare", Information Science Reference IGI Global 2009, ISBN: 978-1-60566-374-6, pages 514-530

Semantic Integration for Science Domains

- Concept of describing scientific domains for *in-silico* experimentation and collaboration
- Based on separation of the *domain* model from the *integration* model
- Facets defined in integration model are automatically mixed-in concepts from domain model
- Proposed, designed and deployed the method for 3 domains:
  - Computational chemistry
  - Sensor processing in an Early Warning System
  - Processing of results of massive bioinformatic computations for protein folding

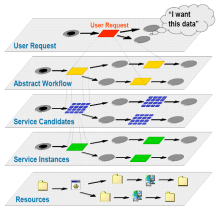




T. Gubala, "Domain Semantic Information for Building e-Science Applications", PhD Thesis, in preparation, promotor M. Bubak, co-promotor P.M.A. Sloot

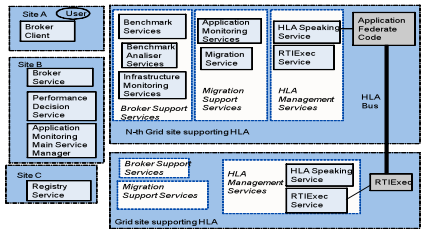
Semantic Workflow Composition

- Dynamic refinement of workflo based on semantic description
- Abstract, functional blocks translated into computation unit candidates
- Concurrency constructs based o Petri Nets
- Runtime refinement: unknown failed branches are re-construct with different computation unit candidates



T. Gubala, D. Harezlak, M. Bubak, M. Malawski, "Semantic Composition of Scientific Workflows Based on the Petri Nets Formalism", in "The 2nd IEEE International Conference on e-Science and Grid Computing", IEEE Computer Society Press, http://doi.ieeecomputersociety.org/10.1109/E-SCIENCE.2006.127, 2006

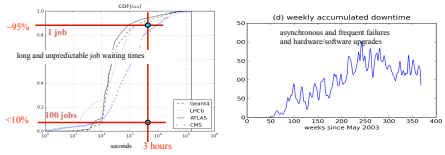
Grid based HLA Simulation Support



K. Rycerz : Grid-based HLA Simulation Support. PhD thesis, University of Amsterdam, June, promoter P.M.A. Sloot, co-promoter: M. Bubak (2006).
K. Rycerz, A.Tirado-Ramos, A.Gualandris, S.Portegies Zwart, M.Bubak, and P.M.A. Sloot. Interactive N-Body simulations on the Grid: HLA versus MPI. International Journal of High Performance

Spatial and temporal dynamics in grids

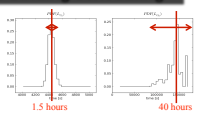
- Grids increase research capabilities for science
- Large-scale federation of computing and storage resources
- 300 sites, 60 countries, 200 Virtual Organizations
- 10^5 CPUs, 20 PB data storage, 10^5 jobs daily
- However operational and runtime dynamics have a negative on reliability and efficiency



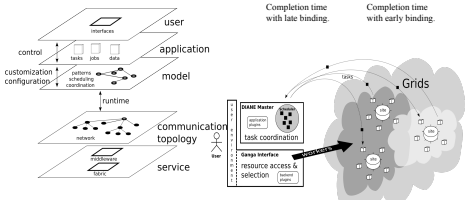
J.T.Moscicki, *Understanding and mastering dynamics in Computing Grids,* UvA PhD thesis, Promotor: M. Bubak, copromotor: P. Sloot; 12.04.2011, at 12.00

User-level Overlay with late binding scheduling

- Improved job execution characteristics
- HTC-HPC Interoperability
- Heuristic resource selection
- Application aware task scheduling



Completion time with late binding. Completion time with early binding.



J.T.Moscicki, M.Lamanna, M.Bubak, P.M.A.Sloot: Processing moldable tasks on the Grid: late job binding with lightweight user-level overlay, FGCS, 2011, in print

Summary

- Combining formal representations of domain knowledge with techniques like data integration, semantic annotations and shared vocabularies, enables the development of collaborative PSE and applications
- A new method of collaborative application development
- Abstract layers to hide technological changes
- Semantic description of applications and resources
- Integration of provenance recording and tracking
- Deployment on available Grid systems, clusters, and single CEs
- Integration of WS, WSRF, components, jobs

Scripts vs workflow applications

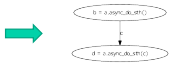
Analysis of script applications

- Applications are developed in Ruby
- Ruby statements are examined to extract information about grid objects, gird operations

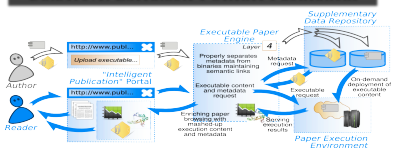
Building workflows

- Is based on data collected in analyzing process
- Grid operations (activities as ellipses, dependencies as arrows

```
a = Gobj.create
b = a.async_do_sth("")
c = b.get_result
d = a.async_do_sth(c)
e = d.get_result
```



DataTrove: Executable e-Science Publications



- A common description schema for primary data (experimental data, algorithms, software, workflows, scripts) as part of publications; deployment mechanisms for on-demand reenactment of experiments in e-Science.
- An integrated architecture for storing, annotating, publishing, referencing and reusing primary data sources.

Challenges

- To go beyond today's social-oriented Web spaces by development of Web-based tools for managing laboratories, running experiments, gathering results, refining methods and achieving scientific goals within virtual groups,
- A model and formalism for specifying interaction schemes and a format for public description of component interaction schemes

Contact person: Marian Bubak & Adam Belloum
Email : M.T.Bubak@uva.nl, A.S.Z.Belloum@uva.nl
Phone: +31 20 525 75 14

COMMIT — VPH-Share — gSLM — MAPPER — Urban Flood — VIROLAB — K-Wf Grid — vl-e