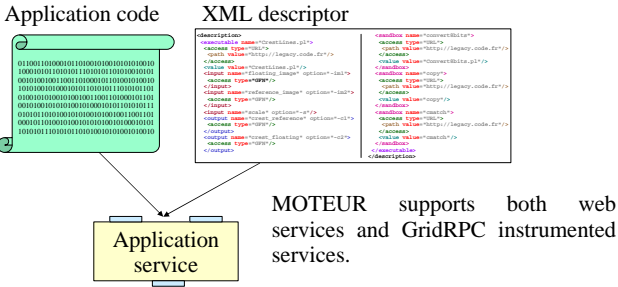


Tristan Glatard
 Johan Montagnat
 Xavier Pennec
 CNRS, I3S laboratory (<http://www.i3s.unice.fr>)
 INRIA, Asclepios team (<http://www-sop.inria.fr/asclepios>)

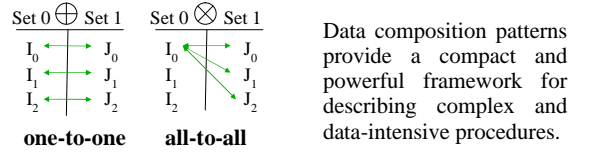
Objectives

Building complex applications by reusing and assembling scientific codes on a production grid infrastructure is greatly facilitated by the MOTEUR workflow manager. MOTEUR adopts the service-based approach which provides maximal flexibility. It eases the description and the enactment of data-intensive applications through a compact and powerful data flows description framework and an optimized execution engine which exploits grid parallelism transparently from the application point of view. In addition, MOTEUR provides a generic service wrapper for embedding non-service aware application code and multi-grids submission services. A service factory enables dynamic application service generation and further optimization through multiple services grouping for lowering grid jobs submission overheads.

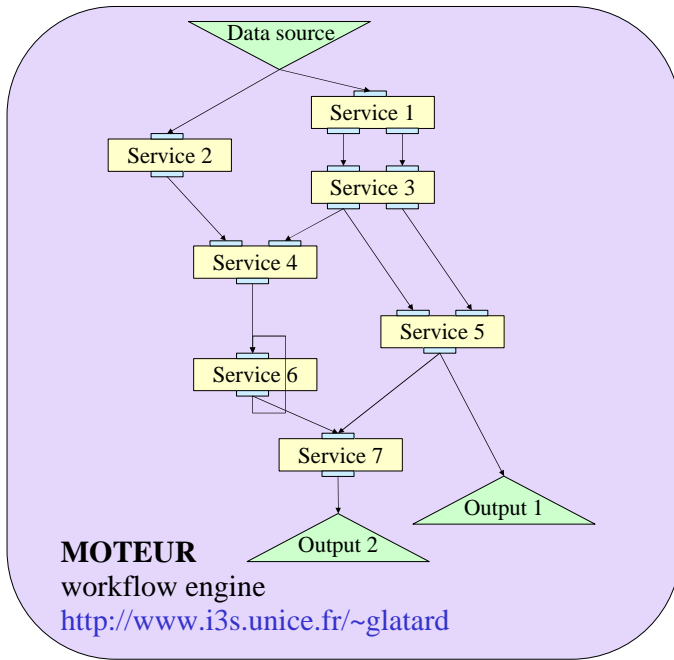
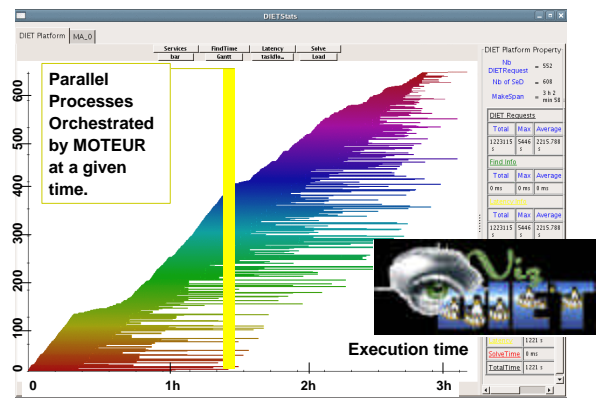
Service-based application code enactment



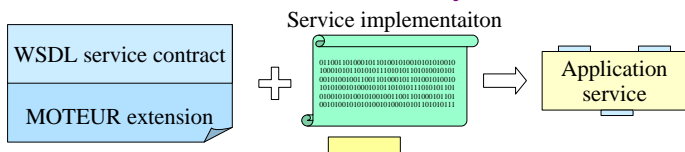
Data composition patterns



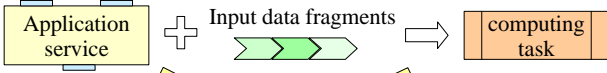
Application execution trace example



Service factory



Submission service



Grid infrastructure execution

11 clusters
2000 CPUs
OAR batch scheduler
Fully reconfigurable

Experimental grid infrastructure

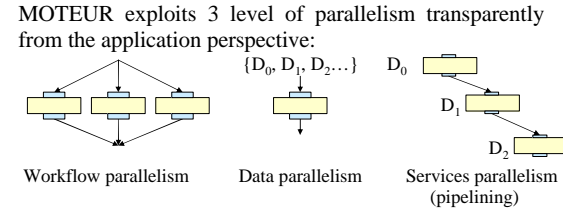
~180 computing centers
~18000 CPUs
~5 PB storage
LCG2 / gLite middleware

Production grid infrastructure

Application targets

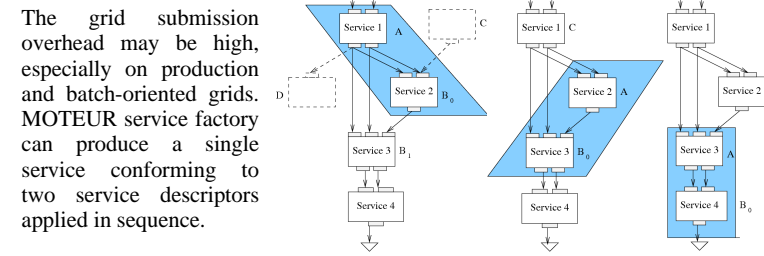
- MOTEUR targets scientific applications. It provides a support for:
- Processing large and dynamic data sets.
 - Parameterizing and updating application workflow.
 - Reproducing experiments.
 - Transparently providing data and code parallelism.
 - Transparently exploiting grid resources.

Transparently exploiting grid parallelism



A complete data processing graph is constructed to prevent causality problems. It also provides data provenance.

Job grouping strategy



- Two sequential services may be grouped provided that:
- Both services provide a MOTEUR extension descriptor.
 - Service grouping does not inhibit any level of parallelism.
- The rule is applied recursively over the complete workflow.

References

T. Glatard, J. Montagnat, X. Pennec. "Efficient services composition for grid-enabled data-intensive applications". IEEE International Symposium on High Performance Distributed Computing (HPDC'06), Paris, France, June 2006.

J. Montagnat, T. Glatard, D. Lingrand. "Data composition patterns in service-based workflows". Workshop on Workflows in Support of Large-Scale Science (WORKS'06), Paris, France, June 2006.

T. Glatard, D. Emsellem, J. Montagnat. "Generic web service wrapper for efficient embedding of legacy codes in service-based workflows". Grid-Enabling Legacy Applications and Supporting End Users Workshop (GELA'06), Paris, France, June 2006.

T. Glatard, J. Montagnat, X. Pennec. "Medical image registration algorithms assesment: Bronze Standard application enactment on grids using the MOTEUR workflow engine". HealthGrid conference, Valencia, Spain, June 2006.

T. Glatard, J. Montagnat, X. Pennec, D. Emsellem, D. Lingrand. "MOTEUR: a data-intensive service-based workflow manager". Research Report I3S, number I3S/RR-2006-07-FR, Sophia Antipolis (France), Mars 2006.

Highlights

- Scientific app.:** Data-intensive Data composition
- Service-based:** Web Services GridRPC
- Performances:** Parallelism Jobs grouping
- Accessibility:** Service wrapper Data provenance
- Grid-enabled:** EGEE Grid5000