Exploiting production grid infrastructures for medical images analysis

Johan Montagnat[1], Tristan Glatard[12], Diane Lingrand[1], Romain Texier[1]

[1] CNRS, I3S laboratory, http://www.i3s.unice.fr/~johan

[2] INRIA Sophia, Asclepios, http://www-sop.inria.fr/asclepios

Introduction

Computerized medical image analysis procedures are increasingly needed for modeling, analyzing and quantifying medical data in order to assist diagnosis, assess pathologies evolutions, plan therapy and train therapists. Such procedures are often complex to set up for different reasons among which: (i) the procedures involve the use of various medical image analysis algorithms which are developed and disseminated over a wide net of research institutions and companies; (ii) they are often complex workflows involving the inter-operation of many of these algorithms; (iii) some procedures may be very demanding in terms of computing power, either to enable complex anatomical and physiological models, or to process large data sets; (iv) the data sources are wide spread over a large net of clinical centers; and (v) most of the data is confidential.

Production grid infrastructures

Scientific grid infrastructures are tackling several of the problems related to medical image analysis in particular. Grids are distributed computing infrastructures. The hardware is a set of standard computers distributed over a wide area network and a software layer, known as *middleware*, is proposing high level services for enabling efficient access to such a system (hiding as much as possible the distributed and heterogeneous nature of the infrastructure to the end user). Grids have evolved over one decade and they have reached a maturity level such that several production infrastructures, such as the world-wide EGEE infrastructure[1], are now deployed. A production system is a 24/7 operational system available to scientists. It is a cross-organizational and multicommunity shared infrastructure (with 18000 CPUs distributed over 180 computing centers in 3 continents, EGEE is servicing tens of different user communities). Cross-organizational user groups, known as *Virtual Organizations*, are gathering users from a same community and controlling their access right to the infrastructure resources.

Grids are intrinsically distributed systems providing vast amounts of computing and storage resources. Today's production grids are deploying interfaces to efficiently manage and transfer distributed files, to control the workload of massively parallel applications and to control users access right in flexible and extensible frameworks. High level services for dealing with specific data formats such as the ones used in clinical environment and for easily deploying large scale data analysis procedures are know being developed. Over the past years, grids have been used for production in many applications related to medical image analysis, thus demonstrating the relevance of this approach [2].

Tackling the needs of the medical community

Grids are providing services to ease the development of large scale medical data analysis procedures. However, the deployment of such procedures is often slowed down by constraints related to the data manipulated. In particular, the confidentiality of medical data makes its distribution of a cross-organizational computing infrastructure difficult. In addition, most medical end users developing such medical proce-

[1] Enabling Grids for E-sciencE, http://www.eu-egee.org

dures are non specialists and they need high level interfaces and support to access grid infrastructures transparently in a familiar and targeted environment. We are tackling these two issues, developing high level middleware services interfacing securely to clinical information systems and easing the description and the efficient handling of large data sets analysis over a grid.

Medical Data Manager. To ease the access to medical image sources and the integration of medical data bases into large scale analysis procedures, we are developing a Medical Data Manager (MDM) interfaced with the EGEE grid middleware [3]. It aims at (i) providing access to medical data sources for computing without interfering with clinical practice, (ii) ensuring transparency so that accessing medical data does not require any specific user intervention, and (iii) ensuring a high data protection level to preserve patients privacy.

This service is exploiting the DICOM standard for medical image transfers on the clinical side and the Storage Resource Management (SRM) on grids. It bridges these two standards by translating on the fly grid file read accesses into DICOM transactions. It benefits from the EGEE middleware capability to manage distributed files, thus enabling the federation of many DICOM servers geographically distributed and it provides a unified view of the data archived. It exploits state of the art encryption and fine grain ACL-based mechanisms to ensure both data protection and access control.

The MDM is in a final development phase. It has been successfully demonstrated at the final EGEE review (May 2006) and its client is now distributed as a part of the new EGEE middleware, known as gLite 3.0.

MOTEUR workflow manager. Image analysis procedures often involve the definition of complex application workflows by assembling existing data processing algorithms. Furthermore, the analysis of large data sets is often needed *e.g.* for statistical analysis or epidemiological studies. The service oriented design is a flexible and powerful framework for inter-operating heterogeneous services. However, it is often not taking into account high performance computing needs. We have developed MOTEUR, a grid-interfaced workflow manager able to wrap legacy codes into a service framework [1].

MOTEUR is providing a high level description framework for designing data-intensive applications with complex data composition requirements. It provides an optimized workflow enactor taking advantage of the grid computing and storage resources for efficiently processing data sets. It transparently parallelize workflow-based applications on a grid, exploiting both data and services parallelism.

Conclusions and references

On top of high level services such as the MDM and MOTEUR, we have implemented the *Bronze Standard* a data-intensive statistical procedure for assessing the performance of rigid registration algorithms. It has validated the sub-voxel accuracy of four different algorithms when registering real brain MR images acquired for oncology [1].

Grid infrastructures have been successfully exploited for production in the area of medical image analysis for several years. The emerging high level services dedicated to medical data manipulation ease the migration of clinical applications. Although today mainly used for scientific research purposes, grids are expected to progressively become a natural tool for federating and processing distributed medical data in clinical context.

[1] T. Glatard, J. Montagnat, and X. Pennec. Efficient services composition for grid-enabled data-intensive applications. In *IEEE Itnl Symp. on High Perf. Distributed Computing (HPDC'06)*, Paris, June 2006.

[2] J. Montagnat, F. Bellet *et al.* Medical images simulation, storage, and processing on the European DataGrid testbed. *Jnl of Grid Computing (JGC)*, 2(4):387–400, Dec. 2004.

[3] J. Montagnat, D. Jouvenot *et al.* Bridging clinical information systems and grid middleware: a Medical Data Manager. In *HealthGrid conference (HealthGrid'06)*, pages 14–24, Spain, June 2006.