Jornada de Seguimiento de Proyectos, 2005
Programa Nacional de Tecnologías Informáticas

GRID-IT: Investigación y Desarrollo de Servicios Grid - Aplicación a Modelos Cliente Servidor, Colaborativos y de Alta Productividad. TIC2003-01318

Vicente Hernández García *
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia

Abstract

GRID-IT focus on easing the migration of applications to a Grid environment by means of the development of a set of components and services. These high-level tools will be extracted from a selection of applications that will be deployed during the project. The applications have been selected taking into account the experience of the research group involved in the project and the model of interaction of the software components. The project aims at covering the whole process of Grid application development, starting from the set up of a testbed, the development of application-oriented components and services on top of standard Grid middlewares and the implementation, validation and study of possible business models for several relevant applications.
Keywords: Grid Services, Components and Applications, DataGrids, High Throughput Computing, Client-Server.

1 Project objectives

The objective of the GRID-IT project is to develop a set of generic Grid services and components on the basis of the migration and implementation of a group of applications. The development considers different paradigms of interaction in a Grid environment. This set of components and services will form a framework to ease the further development of other Grid-based applications in different scientific and technological fields. Other objectives covered in the project include the deployment of a Grid infrastructure, the testing of state-of-the-art middlewares and the validation and study of business models of the developed applications. The schema shown in Figure 1 reflects the structure of the developments performed in the framework of the GRID-IT project. The different applications form the top level layer, under which the GRID-IT components (inferred from the applications) are integrated. These components

*Email: vhernand@dsic.upv.es

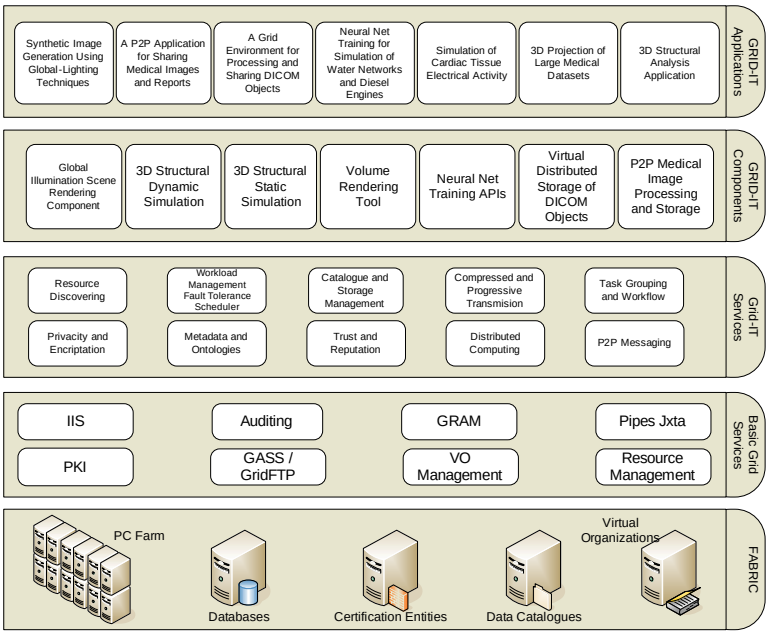

Figure 1: Structure of the GRID-IT project.

are supported by high-level services (the GRID-IT services), which use, in turn, a set of several services provided by the middleware. In some cases these GRID-IT services are re-engineered basic components of existing middlewares.

Some concepts have changed from the start of the project due to the natural evolution of Grid technologies during the last years, leading to an evolution of the project itself. As an example, the definition of OGSI (Open Grid Service Infrastructure), proposed in GT3 [1], has been deprecated and substituted by the specifications defined by the WSRF (Web Service Resource Framework) in GT4 [2, 3]. In the same way, the paradigms of interaction proposed initially (Client-Server, Collaborative and High Throughput) were reconsidered under this situation and the concept of the collaborative model, as it was defined originally, is nearest to the current concept of Datagrid.

The applications that are being developed under each one of the proposed models are the following:

1. **Client-Server Model**: This model includes those applications where the client (commonly a GUI) is a separate application from the server and the clients request from the server most of the information that they need. Two applications have been identified:

   (a) 3D structural analysis.

   (b) Synthetic image generation using of global-lighting techniques (radiosity).

2. **Datagrid Model**: This model comprises a wide set of applications that share data and provide a workbench for the execution of small distributed applications. Two activities have been identified:

   (a) A Peer-to-peer application for sharing medical images and reports.
   (b) A Grid environment for processing and sharing DICOM objects.

3. **High Throughput Computing Model**: This model implies the execution of large sets of experiments on distributed computing nodes on the Grid. Efficiency, fault tolerance and robustness are the main requirements. The applications identified are:

   (a) Neural nets training for the simulation of water networks and diesel engines.
   (b) 3D projection of large medical datasets.
   (c) Simulation of cardiac tissue electrical activity.

2 Success Level Achieved in the Project

In the last two years, Europe has become the driving force for the evolution of Grid technologies through different initiatives, such as the UK e-Science Programme [5], some priorities in the Spanish R&D National Programme, or the same VI Framework Programme where the development of Grid technologies has been promoted by means of several important projects.

This trend and the synergy between GRID-IT and other national and international Grid projects where our research group is participating, have intensively promoted the development of the project and fostered the achievement of its objectives.

Figure 2 provides a quick view of the present status of the project. During the first 18 months the following milestones has been covered:

1. A Grid infrastructure has been deployed. At the moment, the operative middlewares are the following: Globus 2.4, Globus 3.9.5 (beta version of Globus 4), LCG 2.4.0 (Large Hadron Collider Computing Grid) and InnerGrid Nitya. The Grid fabric counts with approximately 350 processors, that include computers dedicated exclusively to the project, non-dedicated desktop computers and shared computing resources. During the second stage of the project, other middlewares oriented to services will be deployed, such as Globus 4 or gLite, an evolution of the LCG middleware developed in the Enabling Grids for E-sciencE project (EGEE [4]) where the research group is participating in the deployment of biomedical applications.

2. The design of the different pilot applications has been completed. Some applications have been deployed using Globus 2.4, and the migration to service-based middlewares, like Globus 4 or gLite, is in progress.

3. The GRID-IT components and services that will form the GRID-IT middleware have been identified. The GRID-IT services are those tools with a more general and horizontal conception which are applicable to a wide range of applications. On the other side, the GRID-IT components are modules that implement some functionalities common to a set of applications in different scientific and technological fields. The services and components identified, extrapolated from the design of the applications, are the following:

   (a) GRID-IT Services:
      i. Work Management:

   A. Workload management, fault tolerance and scheduler.
   B. Task grouping and workflow.
   C. Resource discovering (including multiprocessors).
  ii. Data Management:
   A. Catalogue and storage management.
   B. Metadata and ontologies.
   C. Privacy and encryption.
   D. Compressed and fault tolerant progressive transmission.
  iii. P2P Management:
   A. Trust and reputation.
   B. P2P messaging.
   C. Distributed computing.

(b) GRID-IT Components

  i. 3D structural static simulation component.
  ii. 3D structural dynamic simulation component.
  iii. Scene rendering using global illumination methods.
  iv. P2P medical image processing and storage.
  v. Virtual distributed storage of DICOM objects.
  vi. Neural net training APIs.
  vii. Volume rendering tool.

4. The design of the interfaces of the different services and components has been completed. The integration of the components and services in the middleware and applications have been clearly defined and the interfaces and the functionality described.

5. The implementation of the selected generic Grid services has started. Several services are already available in beta or preliminary versions, such as the Workload Management and the cataloguing and transmission functionalities in Data Management. Those components will be improved in the second part of the project, especially in terms of robustness and efficiency.

The next subsection describes the different applications and components developed in the framework of the GRID-IT project. More information on the services can be obtained in the bibliography or contacting directly the project manager.

2.1 GRID-IT Applications and Components

This section describes the applications which are guiding the development of the project and the components stemmed from these applications. The applications are the base for the extrapolation of the components and services that conforms the main contribution of the GRID-IT project. The components are high-level objects that could be used for a wide set of applications. Not all the applications in the project have contributed in the same way to the GRID-IT middleware. Some of them have produced components or basic services and others both of them.

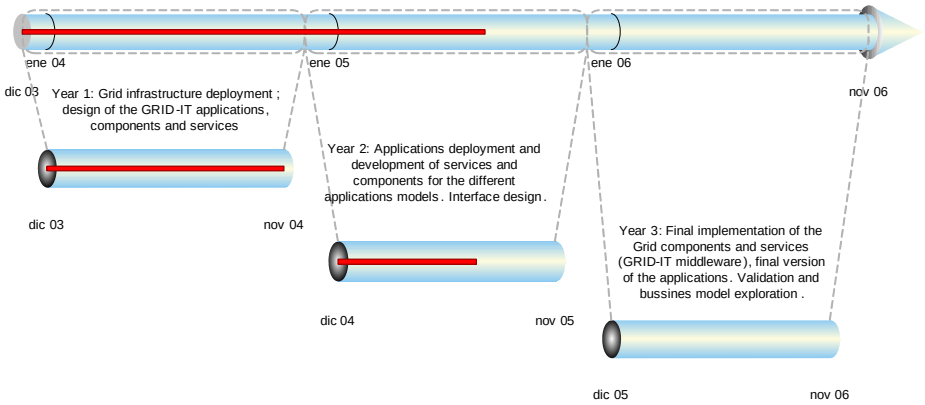

Figure 2: Planification and status of the GRID-IT project

1. **3D structural analysis application**: The objective of this application is to perform the complete 3D structural simulation of a building. The current application is composed by a Grid structural calculation component that receives the structural data to be analysed (from an authenticated user) and using several of the basic services (GRID-IT services and middleware services) performs the simulations under the best conditions. The access to the application is made through a graphic interface that enables the user to access the Grid infrastructure transparently.

   - Generic component associated: Parallel computing-based static and dynamic 3D structural simulation components. These components perform the structural analysis of a building model under different load and sismic hypotheses.

2. **Synthetic image generation using global-lighting techniques (radiosity)**: This application is included within a more general application of electronic commerce. The objective is to generate highly realistic synthetic images in a minimum time. The initial application used CORBA protocols to implement the communication with a cluster of PCs (where the parallel algorithm was running). In the framework of the GRID-IT project this application has been migrated into a Grid environment to benefit from features such as uniformity, transparency, reliability, ubiquity, security, etc. The core of the application is a rendering component that will choose the most suitable resources for the generation of the scene taking into account parameters such as the availability of resources, the priority of the job and the like.

   - Generic component associated: Global illumination scene rendering component. This Grid component allows the generation of a VRML scene (using different algorithms) from a given geometric data.

3. **A Peer-to-peer application for sharing medical images and reports**: This system follows a P2P architecture using the JXTA middleware as basic platform. The system publishes and searches for anonimized information, through a simple interface in a secure

virtual community of medical professionals. The search could be done by key words in the DICOM headers, using tools for content-like searching or by a framework of metadata associated to the shared information. Moreover, the client incorporates some basic processing tools (zoom, panning, thresholding, etc.). The system is totally extendable to other communities where not only could the information be shared but also there would be the possibility of performing lightweight local processes. The application uses some basic services like those related with the compression and transference of data and other high level services.

- Generic component associated: P2P medical image storage and process. This component provides some tools to develop P2P collaborative applications such as distributed searching, eigen trust evaluation of a virtual community, security and privacy, etc.

4. **A Grid environment for processing and sharing DICOM objects**: This application consists in a Grid middleware that implements some components and services for the storage, processing and management of DICOM objects (images, structured documents, signals, etc.). The objective is to create a huge virtual repository of heterogeneous DICOM relevant to a community of users, with the possibility of associating structured meta-information data for performing searches through ontologies frameworks. The application makes use of some basic services related with data management. The architecture of the application is easily extendible to other problems even in other fields different from the medical one.

   - Generic component associated: DICOM objects virtual storage component. This component defines a set of clients and servers that provide the needed tools for the creation of ontologies, virtual repositories, DICOM object sharing and tools for performing intelligent searches through the ontologies.

5. **Neural nets training for the simulation of water distribution and diesel engines**: Some problems, like the simulation of water networks or the contaminants emission of diesel engines, can be simulated by using Neural Nets. This approach, opposite to the use of physic models, is very useful when the computational time is an important factor, e.g. in real-time control. However a model based in Neural Nets must be correctly trained to obtain valuable results. The application developed under the GRID-IT project enables training a high number of Neural Nets increasing the productivity.

   - Generic component associated: Neural Net multi-training component. This component aims to perform a set of Neural Net training tasks, taking into account different configuration parameters.

6. **3D projection of large medical datasets**: The objective of this application is to perform the volumetric reconstruction of CT (Computed Tomography) and MR (Magnetic Resonance) medical images. The system allows taking advantage of the distributed memory and the distributed computing capacity by reconstructing small pieces of the whole image in the distributed resources of the Grid. The generic rendering component of the application splits the whole data volume into small parts and distributes this information among the available resources.

- Generic component associated: Volume rendering component. This component provides the rendering of large image datasets by splitting the data in small pieces and processing this sub-volumes in the Grid nodes.

7. **Simulation of cardiac tissue electrical activity**: The simulation of the electrical activity of the heart is a process with high computational requirements. Typically, a case of study is composed by a set of parametric simulations completely independent that could be executed concurrently. The application, currently developed under GT2.4, allows to perform multiple parallel executions of a given problem by using some of the basic services developed in the framework of the project.

3 Results Indicators

The GRID-IT project has catalysed a very important scientific production. We have to emphasize the considerable number of publications related with the results obtained during the first 18 months of the project (see references [6] to [33]). Moreover, three research reports based on the topics of the project (inside the Parallel and Distributed Computing PhD programme of the Universidad Politécnica de Valencia) have been presented by three researchers of the project to obtain the Advanced Studies Degree (DEA) in Parallel and Distributed Computing.

Apart from these scientific results, derived directly from the project results, other related effects have been motivated by the course of the project. In this sense the research group is involved in several national and international initiatives like the coordination of the creation of the **Metacentre of Supercomputing and Grid Technologies Applications of the Valencian Community** (Universidad Politécnica de Valencia - CSIC - Universitat de Valencia), the participation in the PROFIT TIC project **gCitizen - Development of a Grid Middleware for the Open Management of the Administrative Information and Procedures in the Field of the Public Administration** (FIT-350101-2004-54), in the **Collaborative Science Portal** of Sun Microsystems and Fundación Telefónica, in the **EGEE[4]** project, etc. In the last call of the IST priority of the VI FP, the group has participated in a total of 3 proposals related with Grid technologies, coordinating one of them (an IP proposal called HealinG). At this moment, the group is involved in the preparation of several proposals also related with Grid technologies.

References

[1] *Open Grid Services Infrastructure (OGSI) Version 1.0. Global Grid Forum.* GTuecke, S., Czajkowski, K., Foster, I., Frey, J., Graham, S., Kesselman, C, Maguire T., Sandholm, T., Snelling, D., and Vanderbilt, P. June 2003.

[2] *From Open Grid Services Infrastructure to WSResource Framework: Refactoring & Evolution.* Karl Czajkowski, Don Ferguson, Ian Foster, Jeff Frey, Steve Graham, Tom Maguire, David Snelling, Steve Tuecke. 2004.

[3] *http://www.globus.org/toolkit/docs/4.0/*

[4] *Enabling Grids for E-Science. http://www.eu-egee.org.*

[5] *http://www.rcuk.ac.uk/escience/.*

[6] *A Grid-Enabled Scene Rendering Application.* M. Caballer, V. Hernández y J. E. Román. International Conference on Computational Science 2004 (ICCS 2004). LNCS 3038. (ISBN 3-540-22116-6). Ed. Springer Verlag. 54 - 57. 2004.

[7] *Parallel and Grid Computing in 3D Analysis of Large Dimension Structural Systems.* José M. Alonso, Carlos de Alfonso, Gabriel Garcia, Vicente Hernandez. Euro-Par 2004 Parallel Processing, 10th International Euro-Par Conference. Proceedings Series: Lecture Notes in Computer Science, Vol. 3149. ISBN 3-540-22924-8. 2004.

[8] *Integrating HPC and Grid Computing for 3D Structural Analysis of Large Buildings.* J.M. Alonso, C. Alfonso, G. Garcia and V. Hernandez. Proceedings of the Fourth International Conference on Engineering Computational Technology, Civil-Comp Press, 2004. Paper 92. ISBN 0-948749-98-9. 2004.

[9] *Desarrollo de una Aplicación de Alta Productividad para el Cálculo Estructural de Edificios de Gran Dimensión.* J. M. Alonso, C. de Alfonso, G. García, V. Hernández. Actas de las XV Jornadas de Paralelismo, Universidad de Almería, págs. 277-282. ISBN: 84-8240-714-7. 2004.

[10] *Desarrollo de un Servicio Grid para el Análisis 3D de Estructuras de Edificación.* J. M. Alonso, C. de Alfonso, V. Hernández. Anales de Ingeniería Mecánica. Revista de la Asociación Española de Ingeniería Mecánica. Año 15. Volumen No. 4. ISSN: 0212-5072. ISSN (Edición Digital) 1698 - 5990 Páginas, inicial: 2387 final: 2395. Diciembre 2004. León.

[11] *Minimizando los Tiempos de Análisis 3D de Estructuras de Edificación de Gran Dimensión.* J. M. Alonso, G. García, V. Hernández. Anales de Ingeniería Mecánica. Revista de la Asociación Española de Ingeniería Mecánica. Año 15. Volumen No. 4. ISSN: 0212-5072. ISSN (Edición Digital) 1698-5990. Páginas 2407-2416. Diciembre 2004. León.

[12] *Enabling High Level Access to Grid Computing Services.* J.M. Alonso, V. Hernández, G. Moltó. European Research Consortium for Informatics and Mathematics (ERCIM). ERCIM News. Special: Grids: The Next Generation, ISSN: 0926-4981. Volumen: 59. Páginas: 42-43. Año: 2004.

[13] *Servicios ASP para el Proceso y Transmisión de Imagenes Médicas Tomográficas.* I.Blanquer, V. Hernandez, D. Segrelles. VII Congreso Nacional de Informàtica de la Salud (Inforsalud) ISBN 84-930487-8-X. 2004.

[14] *A Middleware Grid for Storing, Retrieving and Processing DICOM Medical Images.* Ignacio Blanquer, Vicente Hernandez, Ferran Mas, Damià Segrelles. Workshop on Distributed Databases and Processing in Medical Image Computing (DIDAMIC). Working Notes of the Workshop on Distributed Databases and Processing in Medical Image Computing. Rennes (Francia) 2004.

[15] *A P2P Platform for Sharing Radiological Images and Diagnoses.* Ignacio Blanquer, Vicente Hernández, Ferran Mas. DiDaMIC'04 (Rennes, Francia). 2004.

[16] *Empleo de Técnicas Computacionales Avanzadas para la Aceleración de las Simulaciones Cardiacas.* J.M. Alonso, J.M. Ferrero (Jr.), V. Hernández, G. Moltó, M. Monserrat, J. Saiz. Comité organizador del CASEIB 2004. Libro de Actas del XXII Congreso Anual de la Sociedad Española de Ingeniería Biomédica. ISBN 84-688-9318-8. Páginas: 267-270. 2004.

[17] *Globus-Based Grid Computing Simulations of Action Potential Propagation on Cardiac Tissues.* José M. Alonso, Vicente Hernández, Germán Moltó. Proceedings of the Euro-Par 2004 Conference. Lecture Notes in Computer Science. ISBN 3-540-22924-8, ISSN 0302-9743. Volumen: 3149. Páginas: 444-451. Año: 2004.

[18] *Computación de Altas Prestaciones sobre Entornos Grid en la Simulación Eléctrica de Tejidos Cardiacos.* J.M. Alonso, J.M. Ferrero (Jr.), V. Hernández, G. Moltó, M. Monserrat, J. Saiz. Libro de Actas de las XV Jornadas de Paralelismo, ISBN: 84-8240-714-7. Páginas: 271-276. Año: 2004.

[19] *Biomedical and Civil Engineering Experiences Using Grid Computing Technologies.* J.M. Alonso, C. de Alfonso, V. Hernández, G. Moltó. Proceedings of the Parallel Computing (ParCo 2005) Conference. 2005.

[20] *Clinical Decision Support Systems (CDSS) in Grid Environments.* Ignacio Blanquer, Vicente Hernandez, Damià Segrelles, Montserrat Robles, Juan Miguel Garcia, Javier Vicente Robledo. From Grid to HealthGrid IOS Press ISBN 1-58603-510-X. 2005.

[21] *A Grid-based Application for the 3D Structural Dynamic Simulation of High-Rise Buildings.* J. M. Alonso, C. de Alfonso, V. Hernández, G. Moltó. Proceedings of the Tenth International Conference on Civil, Structural and Environmental Engineering Computing. 2005.

[22] *Creating Virtual Storages and Searching DICOM Medical Images through a Grid Middleware based in OGSA.* Ignacio Blanquer, Vicente Hernández, Damià Segrelles. Internacional Symposium on Cluster Computing and Grid (CCGRID05). ISBN 0-7803-9075-X. 2005.

[23] *A Web-Based Application Service Provision Architecture for Enabling High-Performance Image Processing.* Carlos de Alfonso, Ignacio Blanquer, Vicente Hernandez, Damia Segrelles. Lecture Notes in Computer Science, ISBN 3-540-25424-2. ISSN: 0302-9743. 3402, 260-273. 2005.

[24] *Large Medical Datasets on the Grid.* Carlos de Alfonso, Ignacio Blanquer, Vicente Hernández. Methods of Information in Medicine 2005; 44: 172-176. ISSN 0026-1270. 2005.

[25] *The Grid as a Healthcare Provision Tool.* Vicente Hernández, Ignacio Blanquer. Methods of Information in Medicine 2005; 44: 144-148. ISSN 0026-1270. 2005.

[26] *Three-Dimensional Cardiac Electrical Activity Simulation on Cluster and Grid Platforms.* J.M. Alonso, J.M. Ferrero (Jr.), V. Hernández, G. Moltó, M. Monserrat, J. Saiz. Springer-Verlag Lecture Notes in Computer Science. ISBN 3-540-25424-2, ISSN 0302-9743. Volumen: 3402. Páginas: 219-232. Año: 2005.

[27] *An Object-Oriented View of Grid Computing Technologies to Abstract Remote Task Execution.* J.M. Alonso, V. Hernández, G Molto. IEEE CS Press Proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-based Processing. ISBN 0-7695-2280-7, ISSN 1066-6192. Páginas: 235-242. Año: 2005

[28] *A Peer-to-Peer Environment to Share Medical Images and Diagnoses Providing Context-Based Searching.* Ignacio Blanquer, Vicente Hernández, Ferran Mas. 13th Euromicro Conference on Parallel, Distributed and Network-based Processing. Lugano (Suiza). 2005.

[29] *Experiences on a Large Scale Grid Deployment with a Computationally Intensive Biomedical Application.* J.M. Alonso, V. Hernández, G Molto. 18th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2005). Special Track: Grids for Biomedicine and Bioinformatics.

[30] *Computación Paralela y Tecnologías en el Cálculo Dinámico de Estructuras de Edificación* J.M. Alonso, V. Hernández, G Molto. Proceedings of the 12th International Congress on Computer Science Research (CIICC05). Pendiente de aceptación.

[31] *Three-dimensional Structural Dynamic Analysis using Parallel Direct Time Integration Methods.* J.M. Alonso, V. Hernández. Proceedings of the Tenth International Conference on Civil, Structural and Environmental Engineering Computing (CC2005). 2005.

[32] *A Grid-based Application of the 3D Dynamic Analysis of High-Rise Buildings.* J. M. Alonso, C. de Alfonso, V. Hernández, G. Moltó. Proceedings of the The Tenth International Conference on Civil, Structural and Environmental Engineering Computing (CC 2005). 2005. (Aceptado, pendiente de publicación).

[33] *Grid Technology for Structural Analysis.* J. M. Alonso, C. Alfonso, G. García. V. Hernández. Computers & Structures. 2005. (Pendiente de aceptación).