GILDA
Grid INFN Virtual Laboratory for Dissemination Activities

Roberto Barbera[1,2], Leandro N. Ciuffo[1], Emidio Giorgio[1]
Antonio Calanducci[1] and Valeria Ardizzone[1]
[1]*Istituto Nazionale di Fisica Nucleare, Sez. di Catania, Via S. Sofia 64 – 95123, Catania, Italy*
[2]*Dipartimento di Fisica e Astronomia dell'Università di Catania, Via S. Sofia 64 – 95123, Catania, Italy*

Keywords: Grid computing, Training, Virtual Laboratory, gLite, e-infrastructure.

Abstract: The Grid INFN virtual Laboratory for Dissemination Activities (GILDA) is a fully fledged Grid test-bed devoted to training and outreach activities. Open to anyone who wants to have its first hands-on experience with grid systems, GILDA has been adopted as the official training infrastructure by many Grid projects all around the world. All services, tools and materials produced in the past tutorials can be freely used by anyone who wants to learn and teach Grid technology. Additionally, through a set of applications ported on its Grid Infrastructure, developers can identify components and learn by examples how to "gridify" their applications. This work presents the main features of such training infrastructure.

1 INTRODUCTION

Launched in 2004 by the Italian National Institute for Nuclear Physics (INFN), GILDA (the Grid INFN virtual Laboratory for Dissemination Activities) is a fully fledged Grid test-bed devoted to dissemination activities. This infrastructure is open to anyone who wants to have its first hands-on experience with Grid systems. Actually, GILDA can be an important tool for at least three main categories of users:

1. Grid newcomers – people willing to start learning how to use a grid infrastructure;
2. Grid application developers – Just like "webifying" applications to run on a web browser, grid users need to "gridify" their applications to run on a Grid. Through a set of applications ported on GILDA, developers can identify components and learn by examples how to "gridify" their own applications;
3. Tutors – GILDA has been developed keeping training and education in mind. Thus, all services, tools and materials produced in the past tutorials can be freely used by anyone who wants to learn or teach Grid technology.

Indeed, GILDA has been adopted as the official training platform by several former and current Grid projects, such as EGEE/EGEE-II/EGEE-III (www.eu-egee.org), EELA/EELA-2 (www.eu-eela.eu), EU-IndiaGRID (www.euindiagrid.org), EUMEDGRID (www.eumedgrid.org), EUChinaGRID (www.euchinagrid.org), ICEAGE (www.iceage-eu.org), OMII-EU (http://omii-europe.org), BEinGRID (www.beingrid.eu) and many others, becoming a "de facto" standard training-Infrastructure (referred hereafter as t-infrastructure) in Europe and in several other parts of the world for dissemination of Grid Computing technology.

GILDA objectives can be summarized as follows: (i) to raise awareness of Grid Computing benefits; (ii) to provide customized formats for dissemination events, according to the skills of attendants; (iii) to facilitate appropriate free on-line content and services for training purposes; and (iv) to encourage the use of a complete t-infrastructure by new communities.

This article aims at presenting an overview of GILDA facilities as well as to invite the reader to try such t-infrastructure.

2 BACKGROUND

Computational and storage limitations are key issues for organizations that depend on computation-intensive applications. Such organizations are frequently affected by market pressures to reduce deployment time and maintenance costs. Hence,

Barbera R., N. Ciuffo L., Giorgio E., Calanducci A. and Ardizzone V. (2009).
GILDA - Grid INFN Virtual Laboratory for Dissemination Activities.
In *Proceedings of the First International Conference on Computer Supported Education*, pages 273-278
DOI: 10.5220/0001963402730278

they may be looking for ways to improve the effectiveness of their infrastructure or their business processes through transformation, or seeking opportunities for innovation that will benefit the business. Grid is not the answer by itself, but in many of these cases, it can certainly play an important role, allowing immediate productivity and benefits and giving more choices and control on how to purchase and leverage IT power for competitive advantage. The Grid vision is to expand parallel and distributed computation, providing a virtualization of heterogeneous compute and data resources, supporting security policy based resource allocation and prioritization. The Grid is ideal for any applications requiring excellent performance and scalability for their compute-intensive processes (e.g., Monte Carlo simulation, engineering CAD simulations, protein modeling, 3D rendering, computational archaeology investigations, etc.).

In this context, the GILDA Project was born with the aim of offering a one-of-a-kind service for those interested in testing the Grid, using gLite (2008) and the EGEE infrastructure with their own systems. GILDA offers either basic experiences through the "Try the Grid" (EGEE, 2006) walkthrough in minutes or intensive and in-depth training by helping users willing to develop applications to be ported into a Grid environment.

3 DISSEMINATION TOOLS

The main objectives of GILDA activity around the world are to encourage and help new and existing communities to support them for improvement or migration of their applications to Grid infrastructures, to accelerate the adoption of Grid technologies, and to increase the satisfaction of those currently using the Grid services through the communities' feedback. Training activities are a key component of the knowledge dissemination process, ensuring that all users fully understand the characteristics of the offered Grid services and that they have enough expertise to properly use the available Grid infrastructure. In order to achieve the main objectives, several dissemination instruments are used. A brief description of these instruments is presented below.

3.1 Tutorials for Applications Developers

Porting an application into a Grid environment has never been an easy task. In order to enable application developers to get used to the main Grid functionalities, an advanced tutorial has been created fully dedicated to teach them how to "gridify" their applications. Such training events are the perfect scenario to put application experts in tight collaboration with Grid experts.

Hence, the GILDA Team has acquired a good experience in transferring knowledge and know-how to new communities by helping them to integrate their applications into several Grid projects' infrastructures.

3.2 Tutorials for Sysadmin

Tutorials for grid system administrators are organized by the GILDA Team to meet the needs of computer centers interested in joining the GILDA test-bed or other Grid infrastructures. It is also worth mentioning that it is possible to have your own GILDA-like stand-alone Grid installed in your institution. In fact, it is a common practice adopted to support the organization of long training events, such as Grid Schools or graduation courses.

Usually, for a novice user, installing a new Grid service and configuring it properly are not straightforward tasks. In that scenario, a step-by-step guide is desired to help user during the troubleshooting stage. The appropriate documentation on site installation can be found at (GILDA, 2004d).

The sysadmin tutorials are divided into three parts: a theoretical part, a practical one and the hands-on section. In each presentation, the GILDA tutors use slides to show the main steps necessary to install and configure a Grid element. Then, in the hands-on section all participants try to install a Grid element by themselves, assisted by tutors.

Training material is composed of multimedia slides, a Wiki website and a set of "Virtual Services", detailed below.

The most important stage during the preparation of a tutorial is to set up the machines used for the grid node installation: to make it easy, a virtualization technique such as VMware® is adopted in all organized tutorials. In this way, all participants have their own Virtual Machines on which they can work with minimum efforts required from the system administrators.

3.3 GILDA Wiki

The GILDA Wiki (2008) has been created mainly with the purpose of documenting and organizing the huge amount of training material produced so far. Its

contents are freely available on the Web for every interested user.

Moreover, this site is not a simple on-line documentation repository, but an important collaborative tool used by the whole project team. Thus, all registered user can contribute feeding the site with useful additional training material.

The Wiki site is an information source for three target audiences: users, site administrators and application developers. Regarding the users' content, the available material is subdivided into three different complexity levels (Basic, Medium and Advanced), while the site administrators area consists in step-by-step procedures to install and configure Grid services. Finally, the developers section presents the middleware functionalities that can be integrated into the source code using the API.

3.4 Support System

Through a ticket-based support system (GILDA, 2004a), users can send their questions to the GILDA Team and get customized answers to their issues. The GILDA support system is an important communication tool aimed at helping anyone facing problems on using or installing a GILDA site.

3.5 Virtual Services

Through the use of the virtualization technique, it is possible to carry out all Grid elements. This instrument has many benefits, such as increasing services' portability and reducing both the time to put a site in operation and the number of real machines required. The complete list of available virtual services is available at (GILDA, 2004b).

4 THE GILDA STACK

Four main layers can be identified in the GILDA environment, as depicted in Figure 1. At the lowest level, dedicated hardware resources, such as computers and digital networks running the required operating systems and protocols, as well as local resources, which can be either physical or virtual, such as emulations of computers and operating systems on which a Grid is to be built. At the second level resources are made accessible by software services (middleware, databases, APIs, course-specific applications). At the third level, guides and cookbooks suggesting how to build the t-Infrastructure such that everyone can replicate it if he/she wishes or needs it.

Finally, on top there should be matching compositions of: rubrics and instruction texts, example data sets modified in scale and content to suit educational goals and data use policies, exemplar applications, problem sets, software to support exercises, and everything which makes the infrastructure profitable by students.

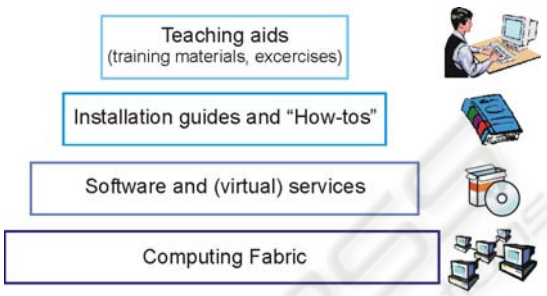

Figure 1: The GILDA stack.

5 THE GILDA TEST-BED

The GILDA test-bed is maintained on a "best-effort" basis. The computational resources usually comes from institutions located in Europe, Asia or Latin America. Therefore new grid sites, using heterogeneous hardware, are frequently joining the t-Infrastructure, just like a "real world" Grid environment.

The test-bed itself is made up of all the components of a larger real Grid infrastructure, including services, resources and monitoring tools. To allow the use of the test-bed, it also features a Virtual Organization (VO) and a real Certification Authority (CA) that grants two-week certificates for the test use of the GILDA infrastructure. In addition, it runs the latest production (and stable) version of the gLite middleware (2008) developed in the context of the EGEE project.

In order to ease the access to its infrastructure, GILDA also provides a grid portal called GENIUS (GILDA,2004c), where different usages are supplied, such as basic content for novice users and full featured ones for more in depth tutorials and demonstrations. At a more advanced level, GILDA offers a one-of-a-kind service for those interested in testing the Grid and gLite with their own software, offering a more intensive and in-depth introduction to the Grid. In a couple of weeks, an user can be trained in Grid usage, his/her software can be modified in order to run on the Grid environment and finally, a GENIUS service can be created to increase the visibility of the work that has been done.

6 VIRTUOUS CYCLE

By its nature, GILDA is one of the key enablers of the "virtuous cycle" to attract and support new communities. Its workflow, depicted in Figure 2, is described a below:

1. Driven either by the dissemination events or scientific papers where GILDA is cited, a novice user can get the feeling of what Grid Computing is and which kind of applications can run on a Grid infrastructure by exploring the GILDA web portal or using the Grid Demonstrator interface (GILDA,2006).
2. Following up the dissemination activities, institutions usually request the organization of Grid training events based on the GILDA t-infrastructure. An interested user, participating in such tutorial or induction course, can go through all the mandatory procedure of the request of a personal digital certificate and subscription to a Virtual Organization and then use the Grid Tutor machine (GILDA,2004c);
3. After participating in a training event or even doing self-training by following the GILDA wiki pages, many users try to port their own applications on the test-bed. In a smaller scale, these users will face all problems of interfacing the Grid services available before entering in a huge production quality e-Science infrastructure;
4. Various applications from different communities ported on GILDA can be either deployed on large e-Infrastructures or incorporated into the Grid Demonstrator, enriching the portfolio of examples that can be demonstrated to new people.

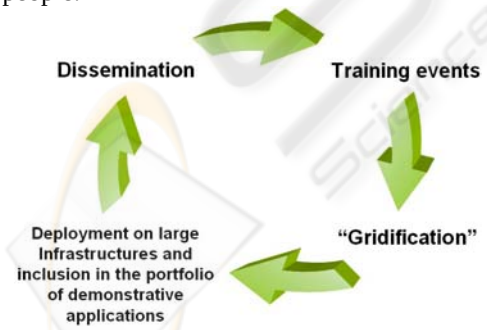

Figure 2: "Virtuous cycle" of applications on GILDA.

7 APPLICATIONS

The Grid environment requires many new skills both for scientists that need to learn how to work on it and for application developers that have to learn how to write and optimize codes to properly interact with the available Grid services and computational resources. As a consequence, disciplinary support is also essential to carry out Grid knowledge management and Grid education, both generic and towards specific application domains. In GILDA a set of "Case Studies" of integrated applications has been created, so that scientist or software developers can explore and compare different approaches for the integration process of their application on a Grid infrastructure.

Currently, GILDA applications portfolio covers a large range of different research domains. A short list of applications successfully ported into GILDA test-bed is presented below.

7.1 GATE

Gate is a C++ platform based on the Monte Carlo Geant4 toolkit (CIRRONE,2005). It has been designed to model nuclear medicine applications, such as PET (Positron Emission Tomography) and SPECT (Single Photon Emission Computed Tomography) in the context of the OpenGATE collaboration (GATE). Its functionalities combined to its ease-of-use make this platform also impressive for radiotherapy and brachytherapy treatment planning.

7.2 Hadron Therapy

HadronTherapy simulates the beam line and particles detectors used in the proton-therapy facility CATANA (Centro di AdroTerapia e Applicazioni Nucleari Avanzate), operational at INFN-LNS. Here, a 62 MeV (62 million of electron volts) proton beam, accelerated by a superconducting cyclotron, is used for the treatment of some kind of eye cancer.

7.3 gMOD

gMOD is a new application offering a content-on-demand service. A user can browse a catalog of multimedia contents and request one to be streamed in real time to his/her workstation using the power of the Grid. Movies can also be chosen by querying an underlying metadata catalog using one or more attributes. The role feature of the Virtual Organization Membership Services (VOMS) is exploited to define normal users and catalog managers. gMOD represents a very interesting use case for the use of Grid both in entertainment and distributed digital content management systems such as the e-learning ones.

7.4 Volcano Sonification

Current knowledge of volcanic eruptions does not yet allow scientists to predict future eruptions. This application represents an attempt to put the scientific community one step closer to the prediction asset by means of the sonification of volcano seismograms. Thus, the translation of the patterns of Mount Etna (Italy) and Mount Tungurahua (Ecuador) volcanic behaviors into sound waves has been carried.
Data sonification is currently used in several fields and for different purposes: science and engineering, education and training. It acts mainly as data analysis and interpretation tool. Figure 3 illustrates its working scheme.

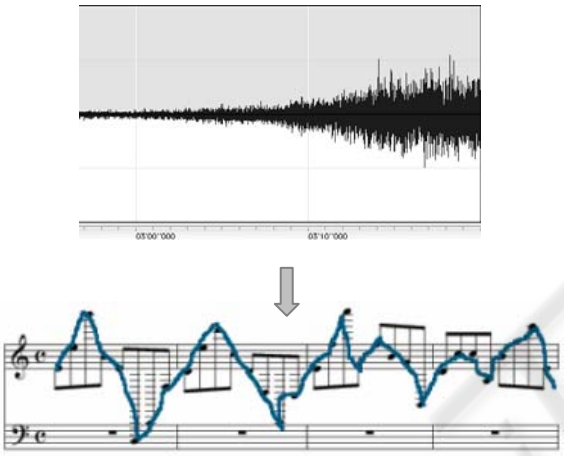

Figure 3: Volcano Sonification – input / output scheme.

8 FORTHCOMING FEATURES

The t-infrastructure provided by GILDA will be continuously used by many EU funded projects, since several project proposals submitted to the 7th Framework Programme calls expressed their intention to use GILDA facilities as part of their training program.

The GILDA hardware expansion plan foreseen the integration of new 64-bits worker nodes by the end of 2008 (up to now, only 32-bits nodes are available). This will lead the availability of new libraries often requested by application developers and the general industry community such as MPI-2.

Additionally, the growing number of users has encouraged the GILDA team to recently announce the incorporation of some new features. These features are:

8.1 gLibrary

gLibrary (CALANDUCCI,2007) an extensible, robust, secure and easy-to-use system to handle digital libraries in a Grid infrastructure. It offers an intuitive Web interface to browse the available entries, and thanks to its powerful “iTunes-like” searching capabilities based on attribute filters, finding a library asset is just a matter of seconds. gLibrary is a flexible system that can be used for different purposes and for different communities that can easily adopt this framework to build their own digital libraries defining types and categories according to their needs.

8.2 Secure Storage Service

This service was carried out by UNICO S.r.l. (http://www.unicosrl.it) in close collaboration with INFN in the context of the TriGrid VL Project (http://www.trigrid.it). This service provides the users with a set of tools for storing confidential data (e.g., medical or financial data) in a secure way and in an encrypted format on the Grid storage elements. The data stored in this way are accessible and readable by authorized users only. Moreover, it solves the insider abuse problem preventing also the administrators of the storage elements from accessing the confidential data in a clear format.

8.3 GRelC

The Grid Database Access Service (GRelC) (2007) aims to provide a set of advanced data grid service to transparently, efficiently and securely manage databases in a Grid environment (FIORE,2007). This framework, developed by University of Lecce – Italy and SPACI Consortium (The Italian Southern Partnership for Advanced Computational Infrastructures), is currently used to support bioinformatics experiments on distributed and huge data banks. The framework provides a uniform access interface to access and interact with heterogeneous DBMS such us PostgreSQL, MySQL, Oracle, DB2, SQLite, etc.

9 CONCLUSIONS

GILDA represents a very valuable tool in the Italian INFN Grid Project (http://grid.infn.it) as well as in several Grid projects around the world. Indeed, its t-infrastructure has been used so far in more than 300 induction courses in 53 different countries all over

the world. Altogether, more than 13000 certificates were issued by the GILDA Certification Authority, and at the time this article was written, more than a thousand users were currently registered with the GILDA Virtual Organization.

Through GILDA a broad range of users are able to get quick and easy access to a "real world" Grid environment. Furthermore, GILDA facilities are open to any organizations or educational institutions who want to adopt it in their Grid training programs. In that sense, the main business benefits of adopting GILDA can be summarized as follows:

1. It is free;
2. No costs to leverage Grid resources are required to use it;
3. It is the first place where a user can takes experience on Grid Computing;
4. It provides an affordable and pleasant experience in Grid training;
5. It provides a good infrastructure where business applications can try "gridifying" operation exploiting the GILDA Team support.
6. It improves local research activity quality providing more power computation and storage resources;
7. It contributes to promote international cooperation and partnership.

ACKNOWLEDGEMENTS

The GILDA Team would like to thank all the partners (project and institutions) that collaborate on proving resources, contents, services and human effort to this project.

REFERENCES

Barbera, R., Falzone A. & Rodolico A. (2003, March). *The GENIUS Grid Portal*. Paper presented at Conference for Computing in High Energy and Nuclear Physics (CHEP 2003), La Jolla, California, USA.

Calanducci,A., Cherubino,C., Ciuffo,L.N. & Scardaci,D. (2007). "*A Digital Library Management System for the Grid*", In: Fourth International Workshop on Emerging Technologies for Next generation GRID (ETNGRID 2007) at 16th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE-2007), GET/INT Paris, France.

Cirrone, G.A.P., Cuttone, G. et al. (2005) *Implementation of a new Monte Carlo-GEANT4 Simulation tool for the development of a proton therapy beam line and verification of the related dose distributions*. In: IEEE Transactions on Nuclear Science, Vol 52, Issue 1, Part 1, pp.262-265.

EGEE. (2006). *Try the Grid*. Retrieved February 8, 2009, from http://egee2.eu-egee.org/try-the-grid

Fiore S., Cafaro M. & Aloisio G. (2007, June) *GRelC DAS: a Grid-DB Access Service for gLite Based Production Grids.* Paper presented at the Fourth International Workshop on Emerging Technologies for Next-generation GRID (ETNGRID 2007), Paris, France.

GATE. (n.d.). *OpenGATE collaboration.* Retrieved February 8, 2009, from http://opengatecollaboration.healthgrid.org

GILDA. (2004a). *Support*. Retrieved February 8, 2009, from http://gilda-support.ct.infn.it/

GILDA. (2004b).*Virtual Services*. Retrieved February 8, 2009, from https://gilda.ct.infn.it/VirtualServices.html

GILDA (2004c) *Grid Tutors*. Retrieved February 8, 2009, from https://gilda.ct.infn.it/grid-tutor.html

GILDA (2004d). *Instructions for sites*. Retrieved February 8, 2009, from https://gilda.ct.infn.it/sites.html

GILDA. (2006). *Grid Demonstrator*. Retrieved February 8, 2009, from https://glite-demo.ct.infn.it/

GILDA. (2008). *TWiki*. Retrieved February 8, 2009, from http://grid.ct.infn.it/twiki/bin/view/GILDA/

gLite 3.0. (2008). *Lightweight Middleware for Grid* Computing. Retrieved February 8, 2009, from http://glite.web.cern.ch/glite/

GRelC. (2007). *Grid Relational Catalog Project.* Retrieved February 8, 2009, from http://grelc.unile.it/

INFN Grid. (n.d.). *INFN Grid - The Italian Grid Infrastructure*. Retrieved February 8, 2009, from http://grid.infn.it/

LCG. (n.d.) LCG Middleware. Retrieved February 8, 2009, from http://lcg.web.cern.ch/LCG/activities/middleware.html