The INFN-Grid Testbed

R. Alfieri a, R. Barbera b,*, P. Belluomo b, A. Cavalli c, R. Cecchini d, A. Chierici c, V. Ciaschini c, L. Dell’Agnello c, F. Donno c, E. Ferro f, A. Forte g, L. Gaido f, A. Ghiselli c, A. Gianoli b, A. Italiano e, S. Lusso e, M. Luvisetto i, P. Mastroserio j, M. Mazzucato k, D. Mura l, M. Reale c, L. Salconi e, G. Sava h, M. Serra m, F. Sataro a, F. Taurino j, n, G. Tortone j, L. Vaccarossa o, M. Verlato k, G. Vita Finzi c

a INFN-Sezione di Parma, Dipartimento di Fisica, Parco Area delle Scienze 7/A, I-43100 Parma, Italy
b INFN-Sezione di Catania, Via S. Sofia 64, I-95123 Catania, Italy
c INFN-CNAF, Viale Berti Pichat 6/2, I-40126 Bologna, Italy
d INFN Sezione di Firenze, Largo E. Fermi 2, I-50125 Firenze, Italy
e INFN Sezione di Pisa, Via Livornese 1291, I-56010 San Piero a Grado (PI), Italy
f INFN-Laboratori Nazionali di Legnaro, Via Romea 4, I-35020 Legnaro (PD), Italy
g INFN-Sezione di Torino, Via Giuria 1, I-10125 Torino, Italy
h INFN-Sezione di Ferrara, Via del Paradiso 12, I-44100 Ferrara, Italy
i INFN-Sezione di Bologna, Via Irnerio 46, I-40126 Bologna, Italy
j INFN-Sezione di Napoli, Università degli Studi “La Sapienza”, P.le A. Moro 5, I-00185 Roma, Italy
k INFN Sezione di Padova, Via Marzolo 8, I-35131 Padova, Italy
l INFN Sezione di Cagliari, Cittadella Universitaria di Monserrato, Strada Provinciale per Sestu Km 0.700, Casella Postale 170, I-09042 Monserrato (CA), Italy
m INFN-Sezione di Roma, Università degli Studi “La Sapienza”, P.le A. Moro 5, I-00185 Roma, Italy
n INFM UdR di Napoli, Complesso Universitario di Monte S. Angelo, Via Cintia, I-80126 Napoli, Italy
o INFN-Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy

Available online 27 November 2003

Abstract

The Italian INFN-Grid Project is committed to set-up, run and manage an unprecedented nation-wide Grid infrastructure. The implementation and use of this INFN-Grid Testbed is presented and discussed. Particular care and attention are devoted to those activities, relevant for the management of the Testbed, carried out by the INFN within international Grid Projects.

© 2003 Elsevier B.V. All rights reserved.

Keywords: INFN-Grid; Grid computing; Testbed; Virtual Organizations

1. Introduction

The rapid growth of the Internet and the current availability of incredibly fast and cheap networks, as fast as the computer’s internal links, can turn into reality the long-standing computer science dream of dissolving a computational machine across the net into a series of special purpose appliances (seen as virtual services) like computing elements, storage elements, meta-data catalogues, resource brokers, information indexes and so on and so forth. The infrastructure of such a “geographic”-size computer is often termed...
as “the Grid”, a distributed computing model where easy access to large distributed computing and data management resources is provided to large mission oriented communities, often referred to as Virtual Organizations (VOs) [1]. Computational and data Grids are, hence, considered as a possible implementation of the concept of co(l)aboratory [2] where scientists and researchers work together to solve complex interdisciplinary problems, despite their geographic and organizational boundaries. According to the so-called “Grid vision” [1], they will just see the Grid as a seamless extension of their own workstation for what concerns both job execution/monitoring and data access/management.

The Italian National Institute of Nuclear Physics (INFN), through the INFN-Grid Project [3], is committed since more than 2 years to set-up, run and manage a nation-wide Grid infrastructure (referred to, in the following, as the “INFN-Grid Testbed” or, more simply, as “the Testbed”) eventually including all the INFN Divisions (24 in total) which are inter-connected by the large-bandwidth Italian Research and Education Network (GARR) [4]. The final goal of the Project is to build a “distributed” computer made of the computational and storage resources available at the various INFN sites continuously accessible by all INFN researchers and associates involved in the INFN Experiments and Projects.

Within this scope and with these objectives, the INFN is one of the main contractors of two major European Union funded Grid Projects such as the DataGrid (EDG) [5] and the DataTAG (EDT) [6] Projects and it is also participating to the CERN LHC Computing Grid Project (LCG) [7], aiming at building up a world-wide multi-level Grid infrastructure operating around-the-clock for the analysis of data coming from the high energy physics experiments planned at the CERN Large Hadron Collider [8].

In this paper the implementation, management and utilization of the distributed INFN-Grid Testbed is presented and discussed within the analogy, mentioned above, of installing and using a single machine. The paper is organized as follows. Section 2 describes the architecture of the Testbed and its layout. The system installation and tests are discussed in Section 3 while authentication and authorization issues are presented in Section 4. Section 5 is dedicated to Grid monitoring and user support while “real life” use cases are presented in Section 6. Lessons learned so far and conclusions are drawn in Sections 7 and 8, respectively. This paper is a resumé of the full internal report, cited as Ref. [9], the interested reader is addressed to for any further detail.

2. The “machine”

At present, the INFN-Grid Testbed holds both sites/machines belonging to the INFN-Grid Projects tests/activities and sites/machines belonging to the Grid infrastructures of the DataGrid, DataTAG and LCG Projects. A partial sketch of the INFN-Grid Testbed, together with the links to the other sites belonging to the above-mentioned European Grid projects is shown in Fig. 1.

In total, the INFN-Grid Testbed consists of a few hundreds of CPUs and more than a Terabyte of disk space “scattered” over 19 INFN Divisions (Bari, Bologna, Cagliari, Catania, CNAF, Ferrara, Genova, Lecce, Legnaro, Milano, Napoli, Padova, Parma, Pavia, Roma1, Roma3, Torino, and Trieste). Machines can be dynamically allocated for test or (pre-)production activities. A typical layout of a Testbed site is shown in Fig. 2. In some sites, different-purposes Resource Brokers and Information Indexes (see Section 3 for the explanation of the nomenclature) have also been configured and are available.

![Fig. 1. Partial view of the INFN-Grid Testbed with its external connections.](image)
3. The installation of Grid services

The so-called Grid middleware installed on the sites belonging to the INFN-Grid Testbed is the one developed within the EU DataGrid Project [10], the INFN is participating to, on top of the Linux Operating System [11].

The installation of the Testbed resources at all sites is carried out by means of Local ConFiGuration system (LCFG) [12], a tool developed at the Department of Informatics of the Edinburgh University and adopted by the Fabric Management Work Package of the DataGrid Project [13] as an ad interim installation and management system. The tool, which is extremely modular, flexible and customizable, is based on a client–server architecture, where a server holds the configuration files for all the various nodes (clients) to be installed. On each client node a daemon gets the specific configuration from the server and updates the node state so that it matches with the node profile stored on the server. The update of a node is done by means of a collection of objects. Each object consists of a script that is responsible for managing a particular software component (e.g., inetd). Thus, there must be an object for each service or application that has to be automatically managed. LCFG provides also the capability of installing a node from scratch. This requires that the software packages and a basic Linux file-system are available via NFS. It also makes use of a DHCP [14] server and an HTTP server. Usually these services are installed on the LCFG server, but it is also possible to use non-dedicated DHCP and NFS servers.

People of the INFN-Grid Project have directly participated in creating the appropriate LCFG profiles for each Grid element. The Grid elements defined so far are the following:

- User Interface (UI): the node used to access the Grid and submit jobs.
- Resource Broker (RB): the core component of the workload management system. Its duty is to find a resource matching the user’s job requirements.
- Logging and Bookkeeping element (LB): the repository for the events occurred in the lifespan of a job.
- Information Index (II): caching information index, based on the GLOBUS MDS-2 [15], directly connected to the RB.
- Computing Element (CE): the gateway to the site nodes. A job dispatched to a CE is then passed to a local scheduler for the execution on the local nodes.
- Worker Node (WN): a single node of a local farm.
- Storage Element (SE): the node providing services to store, locate and replicate the data and to publish information about data availability. It is the interface to both disk and tape servers.
- Replica Catalog (RC): stores information about the physical files on all the SEs.

For the next future a new language (High Level Definition Language, HLDL) [16] has been developed (and is being tested) in order to improve and simplify the definition of these data. Typically, the configuration information is stored in various different text files containing a set of properties describing the node overall configuration. It is possible to combine these files by inclusion and the information contained in these files can be extended and overridden in a similar way as it happens with Object Oriented languages. The set of configuration files is compiled by a tool that produces an individual profile for each fabric node. The output of this process is an XML document containing the entire description of the client node. When the profiles are updated from the source files stored on the LCFG server, a notification is sent to the clients via
the UDP protocol and the profiles are made available to the client via a HTTP server. The client update process is shown in Fig. 3. A detailed description of the LCFG mechanism is described in Ref. [9] while an installation guide is available in Ref. [17].

Since the EDG middle-ware relies on the correct installation and start of several local and network services, a thorough test of the behavior of the Grid Testbed is of utmost importance in all phases of the deployment. In fact, detailed tests must be performed in order to check at least that:

1. all the Grid elements at all sites are configured in the same way for all the supported Virtual Organizations;
2. user’s jobs can be successfully submitted to all Computing Elements which the user is allowed to submit jobs through the appropriate Resource Brokers;
3. the user’s job environment is properly defined at each worker node of each Computing Element not only at run-time but also at compilation and building times.

On top of the evident complication implied in points (1)–(3), there is also the need of automatic tools to test both the middle-ware and the applications deployed on the Testbed due to the usually limited human resources devoted to management tasks and to the large number of sites. In order to cope with the stringent needs described, the INFN-Grid Project has recently formed a testing working group with the goal of setting up a series of procedures (basically shell scripts) to automatically test the various Grid elements at a given site and the basic middle-ware services. The first version of this test suite has been released and it is available for download from [18]. Using these scripts the site managers can:

- verify, for each Grid element (UI, WN, CE, SE, RB, RC, etc.), the correct installation and functionality;
- verify operative interfaces between Grid elements;
- monitor Grid element performances.

The test suite is composed by a set of core functions which are independent of the particular middle-ware release so it can be used with all the next software releases without any changes in the code. The system administrators can run the tests interactively, with the possibility to exit the procedure and fix the problems at their first occurrence, or in batch mode where the results of the tests are written in a detailed report file. In both cases, at the end of each testing phase, a score (number of tests passed/total number of tests) is returned and a threshold can be set on this score to define the level of acceptance of the test. If the result of the test is below the threshold, the system manager is asked to fix immediately the problem(s) or contact the Testbed support. If the threshold is passed but not with full score, the system manager is advised in the actions to take. The tools developed by the INFN-Grid testing group contributed to the effort on tests carried out by the DataGrid TSTG group [19].

After the successful completion of the installation and configuration test suite, reference jobs have been submitted to the whole Grid. The results of the tests can also be stored in a relational database and inspected via a secure web interface. This activity was originally developed within the Grid activities of the ALICE Experiment [20,21] and has recently been extended to the general tests of the DataGrid [22] and INFN-Grid [23] Testbeds. The status of various reference jobs, run on a daily cron-job basis on the Testbed, is recorded in a MySQL database together with the information on the job submission time, the Resource Broker and the Computing Element the
job has been submitted to. Users can then login to the database web interface using both normal HTTP connection and secure connection, select a Resource Broker and a time interval and inspect the statistics of the tests. The system is going to be generalized to send different test jobs for the various Virtual Organizations and/or even more refined procedures to check the interface between the various Grid elements available on the Testbed.

4. Authentication and authorization

Once a site has been correctly configured and properly registered in the INFN-Grid Testbed, it could be used by all allowed users. The authentication and authorization method adopted by the INFN-Grid Project as well as by the European DataGrid Project is based on the GLOBUS Security Infrastructure (GSI) [15] which makes use of asymmetric-key X.509 personal digital certificates issued by trusted Certification Authorities and owned/managed by the Virtual Organizations running on the Testbed into dedicated LDAP servers. According to the GSI architecture the subjects of these certificates are then “mapped” at the various sites onto generic static or dynamic local users’ which will actually “own” the remote jobs while they are executed. The current status of the INFN-Grid Testbed security infrastructure (in common with the DataGrid Testbed) is extensively reviewed in Ref. [24].

A specific contribution of the INFN-Grid Project in this context has been the realization of automatic procedures [9,25] both to manage the VO LDAP servers and periodically download the certificates from them and fill, with or without system manager customizable filters, the files which hold the “mapping” described above.

In order to be included in the LDAP server of the Virtual Organization they belong to, and then be able to use the Testbed, the users must fill and subscribe an agreement based on a clear and well defined policy. The one adopted is described in the “European DataGrid Usage Guidelines” which can be found in Ref. [30].

Sometimes users belong to more than one Virtual Organization. So, for billing or accounting purposes, it is important that they can choose the appropriate local account associated to the proper VO when they are going to execute a job on the Testbed. Since this capacity is not present in the current implementation of the GLOBUS Security Infrastructure, the code has then been modified in order to allow the users to specify a VO during the authentication procedure.

5. Monitoring and user support

Once the Grid elements are installed at the various sites and act as a “single” computing facility at disposal of the users who are allowed to use it, one of the most important issues in managing and actually using the Testbed on a day-by-day basis is to set-up an effective monitoring system. A Grid monitoring system should allow both local system administrators and simple end-users to determine the “health” of the Testbed at any time and follow its evolution in time. They should be able, inquiring the system via, hopefully, a web interface, to keep under control not only the raw resources (CPU load, memory occupancy, services’ availability, disk space, tape status, job manager status, network availability, bandwidth and I/O traffic, etc.) but also sophisticated “high-level” parameters such as the number or running/waiting/failed jobs of a given Virtual Organization or, as another “real life” example, the number of Montecarlo simulated events of a certain type produced so far by a high energy physics experiment. Furthermore, the monitoring system should not provide only historical/statistical information about the status of the Grid infrastructure but also an effective fault discovery, real-time notification and handling system when a problem occurs.

The INFN-Grid Testbed monitoring system is built around NAGIOS [26], a well known scalable and open source network and resource monitoring package with a powerful, friendly and highly configurable web interface. NAGIOS is based on a “plug-in model”; when there is a need to check a service or a host, the system executes a sensor plug-in that performs the check and returns the results to the control logic. Another important feature is represented by the so-called “event handlers”. Event handlers are commands that can be executed whenever a host or a service changes its state. They can be used to pro-actively fix problems before anyone is notified. There are two types of event handlers that can be defined: service event handlers and host event handlers. Moreover, it is possible
to specify global event handlers that should be run for every host or service state change. When a host or a service remains in a critical state and the time specified in the host or service definition has passed since the last notification was sent out, the system sends out a notification to the contact group of people defined for that particular service. Last but not least, the distributed monitoring feature of NAGIOS makes it very appealing to monitor a distributed computing environment like a Grid Testbed in a scalable way. The goal in the distributed monitoring environment is to off-load the overhead and single point of failure when performing service checks, from a central server onto one or more distributed servers in a one-to-one mapping with a Grid information system. The function of a distributed server is to check all the defined services for a set of hosts while the purpose of the central server is simply the listening for service check results from one or more distributed servers and propagate the information to any additional high-level NAGIOS server. For all information about this architecture, see the NAGIOS documentation available in Ref. [26].

The INFN-Grid Project monitoring group has set-up the monitoring web portal Grid Italy Apparatus and Network Observer (GIANO) [27] which wraps a NAGIOS web interface. GIANO monitors hundreds of host and service parameters distributed in several sites and it is currently running since more than 1 year without any failure. As an example, Fig. 4 shows a typical information about the service status of a couple of hosts. Warnings and/or critical conditions are highlighted and when this occurs useful information is automatically reported to the relevant site managers in order to help them fix the detected problems.

A good monitoring system helps a lot users and system administrators to inspect the status of the Testbed and notify them about the occurrence of (possible) problems but gives no support if the problems happen during the installation of the middle-ware or its configuration and use. On this purpose, the INFN-Grid Project has implemented a support infrastructure in order to offer help and answers to various types of requests and questions. Since there are no dedicated human resources to the support activities right now, the service has been implemented on a best effort basis, involving the managers of all the Testbed sites. Currently the service is under test. In order to have an efficient and stable service a dedicated support team is indispensable. The planning of such an operational service for the next year is under discussion. This service is offered to all the members of the INFN-Grid Project community (both site managers and end-users) for installing, configuring and using the Grid. The support team should take advisory input from people.
working on the middle-ware and in the Testbed service groups, and should also collaborate to all the INFN-Grid activities. Users can access the support service through a mailing list (gridsupport@infn.it) or (preferred method) through a web portal system [28] based on PHP-Nuke [29]. Entering the portal users can reach the support resources and get documents through a single point of access. The server also hosts a central repository for documents about the Testbed and keeps track about the problems and their solutions. The support model consists of two steps:

- a user submits a request to the support system and waits for a reply from the support team from other users;
- if the problem has not been solved by the support group then the request is forwarded to other Grid support services and/or mailing lists or to the appropriate middle-ware expert(s).

Fig. 5. Number of Monte Carlo simulated events produced on the Grid by the CMS Experiment. For the details of the figure see text.

Fig. 6. Interactive analysis of Monte Carlo events produced on the Testbed by the ALICE Experiment.
6. “Real life” use cases

The INFN-Grid Testbed has been continuously used by tens of people since 2 years. At the beginning most of the use has just been testing the GLOBUS middle-ware [31] and the first release of DataGrid middle-ware [32,33]. In the last 6 months, on the contrary, “real life” applications have started to be run by all the high energy physics at the CERN LHC the INFN is involved in ALICE [20], ATLAS [34], CMS [35], and LHC-b [36].

Fig. 7. Virtual reality navigation in a geo-physical Digital Elevation Model.

Fig. 8. 3D rendering of the human DNA.
As a quantitative example, Fig. 5 shows the number of Monte Carlo simulated events produced on the Grid in the last few weeks by the CMS Experiment. The simulation of each event takes several hours on a top-class CPU and produces a multi-Megabyte output file. CNAF (first histogram from the bottom) and PD (Padova, third histogram from the bottom) are sites belonging to the INFN-Grid Testbed.

Fig. 6 on the contrary, shows the interactive analysis of Monte Carlo events produced on the INFN-Grid Testbed by the ALICE Experiment.

Although INFN high energy physics experiments are the main “customers” of the INFN-Grid Testbed, it is also open and capable to run other kind of (inter-/multi-disciplinary applications). As interesting examples, Figs. 7 and 8 show, respectively, the virtual reality navigation in a geo-physical Digital Elevation Model, and the 3D rendering of the human DNA, both produced on the Testbed.

7. Lessons learned so far

As shown in previous sections, setting up, operating and managing a nation-wide Grid Testbed and make it to transparently behave a “single” computer for hundreds of users belonging to different Virtual Organization are far from being trivial tasks. The more than 2 years successful experience gained so far at the INFN-Grid Project can be summarized by the following “lessons learned” which may be useful to other people/organizations when planning similar attempts:

- A key issue is the real involvement and continuous commitment of the different sites constituting the Testbed. In this respect, the intrinsic “distributed” but “unitary” nature of the INFN Divisions over the whole Italian Nation is one of the most important causes of success.
- The National Grid has to have a dynamic and self-adapting infrastructure. A good compromise between a rigid management and a complete user freedom is another reason of success. Too much of the latter refrains new users to experiment new tools and applications, while too much of the former makes the Testbed practically un-manageable.
- A real “production Grid” needs a robust and reliable middle-ware. Focus should be put on around-the-clock availability rather than to new fancy, shaky and/or mostly un-used services and this is actually one of the most ambitious goals of the INFN-Grid Project.
- Other key issues in attracting new people to populate the Virtual Organizations and really exploit the Testbed are easiness of use and a continuous, capillary, dissemination activity. In this respect, the development of a user/science web portal (like GENIUS at INFN-Grid) and frequent tutorials and meetings both for end-users and site managers are strategic activities. Effective monitoring systems and virtual help-desks are also of fundamental importance.
- Last, but not least, since Grid technology is a “young” and rapidly evolving field of computer science, an effective age profile of developers, managers and users of a Grid Testbed should be adopted since the beginning together with a good compromise between staff people (who can perpetuate the information and know-how) and non-permanent positions, essential for the technology transfer and the spread of the Grid paradigm outside laboratories and research institutes.

8. Summary and conclusions

The main goal of the INFN-Grid Project is the deployment of a nation-wide Grid infrastructure (a Testbed) capable to efficiently manage and provide effective usage of the large computational and storage facilities located at the various INFN sites inter-connected by the National Research and Education Network. The adoption of the Grid paradigm will eventually induce the seamless integration of all these geographically distributed resources, normally used by people only locally, into a coherent high throughput and availability national computing facility transparently and ubiquitously accessible by all INFN researchers and associates.

In order to achieve this goal a multiple VO oriented Testbed is being set-up and managed and INFN is participating in large international Grid projects in order to improve Grid services functionalities and validate them on large scale international environment together with more effective management tools.
Specific tools have been developed and used to efficiently install and automatically test the Grid middleware on all the resources available. A rather sophisticated user support infrastructure has also been put in place.

The INFN-Grid Testbed is currently used by many people involved in several INFN Experiments for their day-by-day work. Not all problems have been solved yet but we can nevertheless claim that we are going out of the prototype-and-test phase and we are approaching the production phase.

Acknowledgements

The topics and results presented in this document are the outcome of 2 years of intense and profitable collaboration among many people. The authors are very grateful to all the colleagues involved in the INFN-Grid Project and, in particular, to the Testbed site managers and to the Testbed Work Package members for the effort lavished during this period: A. Amoroso, E. Antonioli, S. Barale, M. Biasotto, A. Brunengo, E. Cangiano, M. D’Amato, A. De Salvo, C. De Vecchi, D. Diacono, F. Fanzago, E. Fasanelli, L. Fonti, M. Gambetti, R. Gomezel, T. Macerini, M. Michelotto, D. Rebatto, S. Resconi, C. Rocca, F. Semeria, A. Silvestri, S. Spanu, P. Veronesi and D. Vicinanza. Special thanks are also due to M. Sgaravatto, E. Ronchieri, M. Mezzadri and F. Prelz for their help in managing the work load management services. Thanks also to the DataGrid Project colleagues, especially to the Integration Team Group, for the continuous and helpful cooperation.

References

[3] All information about the INFN-Grid Project can be found at the URL: http://www.infn.it/grid.
[4] All information about CAST, the Italian Research and Education Network, can be found at the URL: http://www.cast.it.
[6] All information about the DataTAG project can be found at the URL: http://www.datatag.org/.
[7] All information about the INFN-Grid Project can be found at the URL: http://www.infn.it/grid.
[8] All information about the ALICE Experiment can be found at the URL: http://www.cern.ch/Alice.
[10] All information about the CMS Experiment can be found at the URL: http://cern.ch/cms.
[13] All information about the ATLAS Experiment can be found at the URL: http://atlas.cern.ch/.
[14] All information about the LHC Computing Grid Project can be found at the URL: http://lhc-cmg.web.cern.ch/LHCgrid/.
[15] All information about the middleware released by the DataGrid Project can be found at the URL: http://marianne.in2p3.fr.
[22] All information about ATLAS can be found at the URL: http://www.cern.ch/Atlas.
[23] All information about the CMS Experiment can be found at the URL: http://cern-ch/cms.
[24] All information about the LHC-b Experiment can be found at the URL: http://www.cern.ch/LHCb.
[25] All information about NAGIOS can be found at the URL: http://www.nagios.org.
[28] Go to http://cvs.infn.it/cgi-bin/cvsweb.cgi/Atk/.
[32] All information about the middle-ware released by the DataGrid Project can be found at the URL: http://marianne.in2p3.fr.
[33] Go to http://cvs.infn.it/cgi-bin/cvsweb/cgi/Atk/.
[34] Go to http://afs.ecn.infn.it/.