Mem. S.A.It. Suppl. Vol. 13, 158

Memorie della
Supplementi

The Grid in INAF

C. Vuerli[1,2], G. Taffoni[1,2], M. Sponza[2], and F. Pasian[1,2]

[1] Istituto Nazionale di Astrofisica – Information Systems Unit
[2] Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Trieste,
Via Tiepolo 11, I-34143 Trieste, Italy
e-mail: vuerli@oats.inaf.it
e-mail: taffoni@oats.inaf.it
e-mail: sponza@oats.inaf.it
e-mail: pasian@oats.inaf.it

Abstract. This paper presents an overview of the Grid-related projects in which Insitutes of INAF (Istituto Nazionale di Astrofisica) were involved, starting from the GRID.IT project until the recent and currently in progress participation to EGEE (Enabling Grids for E-sciencE), the main project for the setup of a Grid Infrastructure for Science in Europe. The paper will give an overview of these activities putting particular emphasis on some key pilot projects, like the simulations of the Planck mission and the development of tools to widen the Grid capabilities to meet the needs of astrophysical applications.

Key words. Grid: applications – Grid: Tools and Services – Grid: Infrastructure and Projects – Grid: Simulations – Grid: Databases and Scientific Instruments

1. GRID.IT and DRACO

GRID.IT was a three-years multi-disciplinary project whose goal was to validate a national Grid infrastructure for scientific applications. The project was funded by the Italian Ministry for Research and Education. INAF participated with three applications and three Institutes: the Astronomical Observatories of Trieste, Padova and Napoli. The three propoesed applications aimed at: 1) enabling grid access to astrophysical databases (archives and catalogues), building on the existing experience in the Astronomical Observatories of Padova and Trieste; 2) porting to the grid of data reduction pipelines, aimed at the processing of large amounts of data coming from new sky surveys, and building on the experience developed in Naples, e.g., with the participation in the ASTRO-WISE consortium; 3) controlling remotely automated telescopes and instrumentation, building on the experience developed in this field at the Astronomical Observatory of Trieste Baruffolo (2007).

The main goal of the **DRACO** project was to trigger within the Italian astrophysical community the process of porting the on the Grid paradigm their own computing-intensive applications, either purely computational, or performing data processing, or massively accessing data. The project plan was organized in order to start from the operational kernel composed of the nodes of the Italian Grid for Research funded by FIRB (Fondo Italiano per la Ricerca di Base), and to ex-

Send offprint requests to: C. Vuerli

tend their number to add functionality, qualified know-how and high-level scientific applications. As a consequence of this, a Grid for astrophysical research, even in the presence of limited investments, would have brought a remarkable improvement to the community in terms of the added value provided by the high-level astrophysical applications, flexibility in the use of application software, ease and transparency in accessing data, and available computing power, allowing Italy to remain scientifically competitive with other national communities. DRACO represented an extension of the GRID.IT project; three new partners joined the three institutes participating to GRID.IT bringing new applications to gridify. The Astronomical Observatories of Rome and Catania were among the new partners Vuerli (2007).

2. Planck simulations in EGEE

Planck, the ESA (European Space Agency) satellite aimed at mapping the microwave sky through two complete sky surveys, will fly in 2008. It is an extremely demanding space mission in terms of computing power and data storage. Planck simulations mimic the whole mission starting from a virtual sky (ideal or contaminated by introducing several noise sources, see fig. 1). Their main goal is the validation of the acquisition and reduction procedure that will be used to build the final scientific data during the operative phase of the mission. The Grid technology seems to be a promising answer to data storage and processing needs of the satellite. In the framework of the EGEE grid infrastructure, we managed to run a number of experiments aimed at designing and defining an application specific environment for the simulation software and data sharing. Our successful experiments demonstrated that the "gridification" of Planck pipelines is not only possible but also extremely convenient in terms of data processing speed and data sharing. Planck simulations is an officially supported pilot application in EGEE since November 2004 when the application was approved by EGAAP (EGEE Grid Applications Advisory Panel). We started

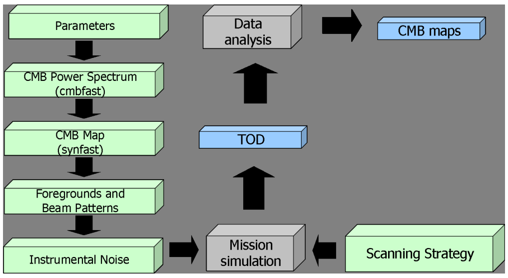

Fig. 1. Schema of Planck simulations

by comparing a partial simulation of the mission run on a dual CPU workstation 2.4 GHz equipped with 2 GB of RAM and on the Grid; we obtained a gain factor of 13 for short simulations (it is assumed in this case that the spacecraft is extremely stable during operations) and of 16 for long simulations that take into account the real situation where the pointing direction of the satellite is not stable due to small drifts. This first results confirmed that the Grid paradigm could be the right answer to the computing needs of the Planck mission so we went ahead in progressively porting the simulation of the whole mission in Grid. After the conclusion of EGEE, the application continued to be supported also in EGEE-II and now it is going to be ported in the forthcoming EGEE-III project. A Planck VO (Virtual Organization), led by INAF Trieste, was set up. All users contributing to the Planck simulations software and located in six different European countries are now registered with this VO and share part of their own home resources within it. More details about the gridification of the software simulating the Planck mission can be found in Taffoni (2007).

3. Recent developments

3.1. The astrophysical cluster in EGEE

The AA (Astronomical and Astrophysical) Cluster in EGEE aims at establishing and consolidating a well motivated astronomical community in Europe that makes use of the Grid technology. The astrophysical cluster in EGEE was created in January 2007 and its main goals are: a) establish and consolidate a well motivated astronomical community in EGEE that makes use of the Grid technology; b) offer to this community a rich variety of hardware/software resources; c) open new opportunities to foster and strengthen scientific collaborations; d) demonstrate that the Grid is more and more helpful when different Grid infrastructures interoperate and different VOs support each other. A number of astronomical Institutes of different European Countries are now part of the cluster; INAF is among them and leads the construction of the cluster and its coordination. The detailed workplan for activities to be carried out within the cluster is not defined yet; it is however under preparation and all participating Institutes will contribute to it. The first preliminary version of it will be brought for approval at the EGEE-III NA4 kick off meeting that will take place in June 2008 (the Network Activity 4 of EGEE is devoted to the User Community Support and Expansion. This includes tasks such as supporting applications and identifying new users). Although the subtasks that will compose the workplan have not been identified yet it is however possible to list a first kernel of activities that will be further detailed:

- *Documentation*. Collection of useful documentation from different sources (especially from NA4) and possibly set up of a specific documentation repository for the cluster.
- *Hardware*. It is extremely important to foster the sharing of new resources within all AA-related VOs by different Institutions and groups of users around Europe that will join the cluster.
- *Software*. The software subtask may be in turn subdivided in:
  - Development of scripts and procedures to make easier the porting of applications in Grid. To achieve this goal the selection of portals suitable to be used for AA applications play an important role.
  - Development of tools and software services that extend the Grid M/W (middleware) and therefore the suite of functionalities offered by the Grid.
  - Census of gridified applications to select good astrophysical demonstrators to be used for training and dissemination purposes.
  - Selection of tools already developed within EGEE and useful for astrophysical applications.
- *Training and dissemination*. The EGEE project already provides training services for all disciplinary clusters, both for Grid novices through introductory training events and for Grid application developers. The AA cluster therefore will take in charge only the organization of training events addressing topics specific of astrophysical applications. Support by EGEE teams will be asked in this case.

It is worth to highlight however that the core activity within the cluster will be the porting of astrophysical applications in EGEE.

3.2. IGI

EGEE is the pilot project of the EU (European Union) for what concern the Grids. Its goal is the creation and consolidation of an European Grid Infrastructure for the scientific community with the perspective of opening this infrastructure also to industry and private companies in the future. Until now EGEE was financed by the EU through the various framework programs. The third edition of EGEE (EGEE-III) starts in May 2008 and lasts two years. An issue concerning the sustainability of the European Grid Infrastructure therefore must be afforded. This infrastructure in fact should continue to live over the termination of the EGEE project.

In this perspective a new entity named EGI (European Grid Infrastructure) is going to be constituted to make this infrastructure stable and self-sustainable; in practice EGI plays for the Grid the same role that GÉANT plays for the physical network infrastructure (http://www.geant.net/). EGI is the entity where all NGIs (National Grid Infrastructures) will join ; IGI (Italian Grid Infrastructure) is therefore the Italian NGI. For the time being IGI is a JRU (Joint Research Unit) within the

EGEE-III project. In the near future IGI will evolve in an autonomous entity to manage the Italian NGI.

The main purpose of IGI is to harmonize the various Grid infrastructures existing in Italy. The proposal for the creation of IGI was formulated by INFN that currently leads the initiative; INAF is one of the founding partners.

4. GDSE and GICS

The following two subsections illustrate the activity carried out in INAF during the last five years to integrate in Grid two important resources for the astronomical community, namely databases and scientific instrumentation. The evolution of the activity aiming at integrating databases and the Grid is quite evident and was supported by several projects. The activity related to the Grid and astronomical instrumentation did not evolve in the same way. The main reason for this is that different alternative solutions were proposed to integrate instruments and the Grid. One of these (the GridCC project mentioned below), in particular, was funded by the EC (European Commission) through the sixth framework program. Although the peculiarity of the solution proposed by INAF (it is completely based on the Grid technology), the competition of other funded projects made extremely complicated the procurement of the necessary resources to undertake the planned activities. In conclusion the activity on Grid and databases in INAF did not evolve as expected due to the lack of economic support.

4.1. Accessing databases through the Grid

The importance of accessing databases via Grid was clear yet at the beginning of the GRID.IT project (section 1). Typical astrophysical applications make intensive use of databases and the Grid middleware in use with the GRID.IT project did not support any mechanism to access a database from the Grid. This represented a serious problem for the porting of astrophysical applications so after having evaluated some tools at that time on the market, it was decided to undertake a development activity on the Grid middleware with the ambitious objective of producing a tool completely built on top of Globus to make the Grid able to access databases as any other resource like a computing element or a storage element. The experiment was completely successful and GDSE (Grid Fata Source Engine) was produced; thereafter a new Grid element specializing a classical CE (Computing Element) and built on top of GDSE was created; this new Grid element is the "QE" (Query Element) allowing Grid users to write extended jobs that combine computation and access requests to one or more databases (see fig. 2). GDSE indeed is general enough to be useful for any application that needs to handle databases via Grid. The first prototype of the GDSE was delevoped through a collaboration between INAF and INFN. This first prototype was subsequently further consolidated and a production release of it was implemented thanks to a 1-year technological transfer project financed by UIT, the Office for technological innovation of INAF. The company that participated to this project as industrial partner was NICE s.r.l., an Italian company that produced the Grid portal named "EnginFrame" (http://www.enginframe.com); thanks to this project GDSE is now fully integrated in this portal so from now on people using EnginFrame can exploit the capabilities offered by this middleware extension. In the framework of the same project, a version of GDSE compatible with gLite,

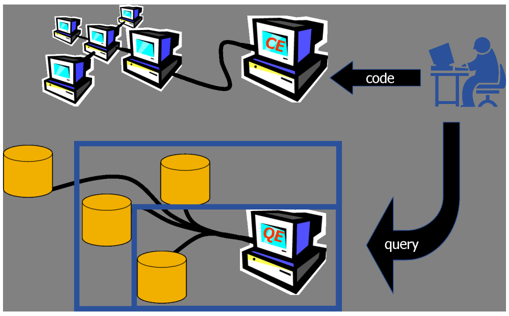

Fig. 2. Query jobs are routed to the QE

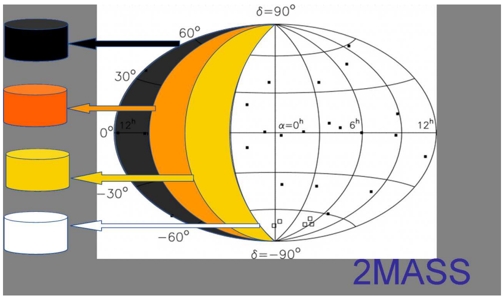

Fig. 3. Split of the 2MASS catalogue

the new M/W supported by EGEE, was also produced. Recently the collaboration between INFN and INAF was resumed to go on with further development on GDSE given the renewed interest that several communities demonstrated for this technology. Through the collaboration between INFN and INAF, GDSE will be soon integrated in CREAM (http://grid.pd.infn.it/cream/). Fig. 3 gives an example of usage of GDSE to integrate an astronomical catalogue in Grid. In this case the GDSE-compatible version of EnginFrame is used to access the 2MASS catalogue. Because of the size of the catalogue, it was split in more parts stored in different databases. In this case the GDSE capability of accessing and routing queries to different databases at the same time is exploited. Detailed information about GDSE can be found in Taffoni (2008). A very important additional note is worth to mention here. Given its capability to integrate databases in Grid, the GDSE plays an important role for what concerns the interoperability between the Grid and the VObs (Virtual Observatory). The interoperability between these technologies is a key factor for astronomers who can rely in this way on fully integrated working envoronments where they find everything they need to run their applications. Further details about interoperability between the Grid and the VObs via GDSE can be found in Taffoni (2006).

4.2. Instrumenting the Grid

As in the case of databases, the problem of intrconnecting the Grid and astronomical instrumentation dates back 2002, when, in the framework of the Italian GRID.IT project (see section 1), the porting in Grid of an application aiming at remotely monitoring astronomical telescopes and related instrumentation was attempted. To carry out this task, a number of tools and services produced by similar projects were evaluated. In particular the GridCC project, based on Web Services was carefully examined (http://www.gridcc.org/cms/). From the point of view of the Grid architecture, the design of the GICS (Grid Instrument Control System) is closely related with what already done for the GDSE. GICS does not represent a custom solution to meet specific astronomical requirements; rather, it was thought and designed having in mind a general-purpose approach based on the concept of "Generalized Grid Element". The GICS can be used in all cases where a generic instrument has to be monitored and controlled via the Grid. When a new instrument has to be interfaced to the Grid a new specialization of the Grid-embedded generic instrument will be produced in a way that the specific problem and kind of instrumentation to be monitored/controlled is taken into account. Once the GICS enabled Grid architecture has been set up, the Grid infrastructure is ready to include the new element "IE" (Instrument Element) of the Grid. As the QE is built on top of GDSE, in the same way the IE is built on top of GICS. The IE is a CE able to handle, monitor and control requests to a specific Instrument Control System via Grid. As for the QE, the IE inherits the whole set of functions of a classical CE. In addition, the IE exploits all new functionalities offered by GICS to handle jobs whose target is the interaction with Instrument Control Systems. Further details can be found in Vuerli (2007).

5. Other Grid related projects in INAF

Besides the Grid projects and activities illustrated in previous sections, it is necessary to mention two important projects with the participation of INAF Institutes that led to the setup of suercomputing facilities and regional Grid infrastructures and to the development of relevant scientific and technological activities by exploiting these infrastructures in the southern

Italian regions. All these projects are funded through the PON (Piano Operativo Nazionale), the plan of the Italian Government to sustain the Scientific and Technological Research in South Italy. The two projects are only mentioned here as they are illustrated in detail in dedicated papers.

- *Cometa* in Sicily with the participation of the Astronomical Observatories of Catania and Palermo.
  http://www.consorzio-cometa.it/
- *CyberSar* in Sardinia with the participation of the Astronomical Observatory of Cagliari.
  http://www.cybersar.com/

6. Conclusions

The startup of Grid-related activities in INAF dates back 2003, with the participation to the project funded by FIRB. Since then INAF was involved in a growing number of projects and activities both at national amd international level. Thanks to these projects, groups of researchers with a deep knowledge of the Grid technology are now present in INAF and can play an important role in training and disseminating this know-how within the Institute. Many users could take advantage of this as it was demonstrated that the Grid can be the right solution for a wide range of astrophysical applications particularly challenging for what concerns the amount of requested resources. Technological expertise range from techniques and best practices to port applications in Grid to the development on the Grid middleware to produce tools and services requested by applications, and so forth.

INAF researchers have now important roles in several projects and led important international groups like the astrophysical cluster and the working group dedicated to the Grid and Databases in EGEE or the leadership role in promoting a stronger interaction between the Grid and the Virtual Observatory with the coordination of the WP5 (Virtual Observatory and the Grid) in the framework of the VO-DCA (Virtual Observatory - Data Centres Alliance) project.

In conclusion, INAF has got a strategic position for what concerns the Grid technology. A more consistent amount of funds however should be allocated in the near future to these activities to avoid the dispersion of the know-how attained in these last years.

References

Taffoni, G., et al. 2006, Grid Data Source Engine: Grid Gateway to the Virtual Observatory, Astronomical Data Analysis Software and Systems XV, Publications of the Astronomical Society of the Pacific, 351, 2006, p. 508

Baruffolo, A., et al. 2007, WP10: Astrophysics in GRID.it, Grid-enabled Astrophysics, L. Benacchio & F. Pasian eds., Polimetrica International Scientific Publisher, 2007, p 13-26

Taffoni, G., et al. 2007, Enabling Grid technologies for the Planck Space Mission, Future Generation Computer Systems, 23, 189-200

Vuerli, C., et al. 2007, Draco Project: Activities, Results and Future Perspectives, Grid-enabled Astrophysics, L. Benacchio & F. Pasian eds., Polimetrica International Scientific Publisher, 2007, p 27-42

Vuerli, C., et al. 2007, Instruments in Grid: The new Instrument Element, 2nd International Workshop on Distributed Cooperative Laboratories: Instrumenting the Grid, 2007

Taffoni, G., et al. 2008, The Query Element: Grid Access to Databases, Grid Computing Research Progress, Jiannong Wong eds., Nova Publishers