Mem. S.A.It. Suppl. Vol. , 1

Memorie della
Supplementi

INAF Archives in the Framework of the Astronomical Data Grid

F. Pasian[1], R. Smareglia[1], C. Vuerli[1], A. Zacchei[1], N. Lama[1], L. Benacchio[2]

[1] INAF/Oss. Astr. di Trieste, Via G.B.Tiepolo 11, I-34131, Trieste
[2] INAF/Oss. Astr. di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova

Abstract. The Grid paradigm for accessing and exploiting network-distributed facilities has been extended to the Data Grid concept, based on storage and processing systems, and metadata and communications management services. Data Grid and the International Virtual Observatory are the two environments the Italian astrophysical community will have to deal with in the future. This paper describes how archives will interface with these new realities, and the pilot Long-Term Archive for the TNG is brought as an example.

Key words. standards – data processing – catalogs – astronomical data bases: miscellaneous – data grid

1. Introduction

A new paradigm for accessing and exploiting network-distributed facilities, namely the Grid, has recently gained importance and momentum; it is common belief that the concept of Grid will be the natural extension of the web, since it allows, besides the passive access to the resources (mainly information) available on the net, an active usage of the resources themselves, e.g. allowing the user to access distributed computing resources. The Data Grid extends this concept, being based on two fundamental services, namely: storage and processing systems, and metadata and communications management. The astrophysical community is implementing this kind of structure by defining, at an international level, the concept of "Virtual Observatory", which relies on the Data Grid paradigm.

2. The Grid – basic principles

GRID is a collaborative set of computing, data storage and network resources belonging to different administrative domains that has knowledge about the status of its components through an active, distributed information service.

Resources can be seen as represented by Logical Machines, which can belong to the following types:

UI: User Interface (hosts all client programs through which the user may interact with the Grid);
RB: Resource Broker;
IS: Information service (deals with the state of the GRID infrastructure);

Send offprint requests to: F. Pasian
Correspondence to: via G.B.Tiepolo 11, I-34131 - Trieste - Italy

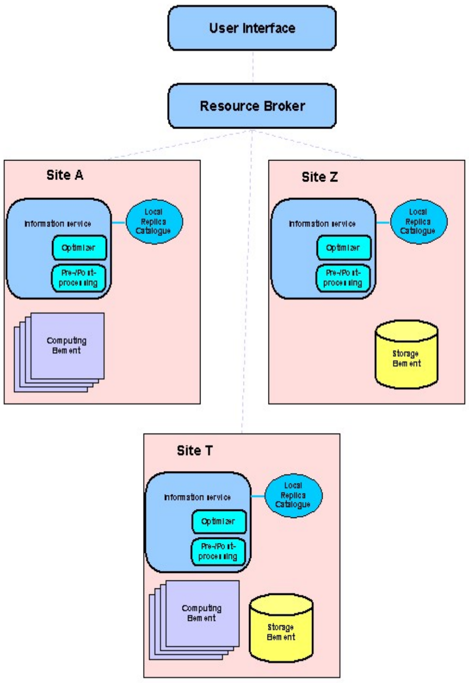

Fig. 1. Scheme of a Grid-enabled system with different types of Logical Machines

CE: Computing Element (supplies CPU power); both Gatekeeper (Front-end Nodes) and Worker Nodes are considered to be CEs;
SE: Storage Element (provides a storage resource, in particular access to mass storage systems);
RC: Replica catalog (manages and maintains the catalog of data set replicas).

In Figure 1 a Grid-enabled system is depicted. Each of the sites composing it can host different Logical Machines, of the types listed above.

3. The International Virtual Observatory

In modern astronomy the need is often felt to match multi-source data, working in a multi-frequency domain, checking information throughout a complete set of data providers; to fulfill such needs is the aim of the International Virtual Observatory (VO). From the technical point of view, the first step of this rather complex mechanism is to be able of handling data spread over different distributed repositories.

The International VO will therefore allow all archives to "speak the same language", i.e.:

- all archives will be searchable and analyzable by the same tools;
- all data sources will be rapidly accessible via a uniform interface;
- all data will be held in distributed databases that must appear as a single one.

The VO experience has been built up from the AstroVirtel project which, under the auspices of OPTICON, has expanded into the Astronomical Virtual Observatory (AVO) project. AVO is a phase A study, supported by six European data providers, for the creation of a prototype multi-wavelength virtual astronomical observatory for the scientific community. All the above projects have been funded by the European Union under the auspices of its Framework Programme no. 5 (FP5). Given the success of the AVO study, a new application is being made to EU FP6 for a full-fledged Euro-VO.

Furthermore, an International VO Alliance (IVOA) has been founded, linking together all VO activities in the world. Members of IVOA are United States, Europe, United Kingdom, France, Germany, Russia, Canada, Australia, China, Japan, Korea and India. Italy, represented by INAF, was the last to join, in March 2003.

4. Interoperability of Italian archives in the VO framework

For the Italian archives of astrophysical data to participate in the VO and Grid activities, it is essential to adhere strictly to the standards defined by the international projects in the field.

A first step in this direction has been made by the prototype TNG Long-Term Archive. One of the aims of its development activity was to provide tools and expertise to create and achieve a high degree of interoperability with other archives at the national and international level. The following steps have been followed:

- use of internationally-defined standards in the implementation of the prototype:
  - FITS (the Flexible Image Transport System),
  - ASU (the Astronomical Server URL, a standard to generate queries for retrieving tabular data and catalogues in astronomy),
  - Astrores (a tool decribing Astronomical Catalogues and Query Results with XML),
  - a preliminary version of VOtable (an extension of Astrores allowing to take binary data into account);
- installation of a name resolver using the SIMBAD facility;
- definition of a higher level of interoperability based on the availability at the CDS of the TNG catalog of observations.

The link of the Italian community with the international projects is actively pursued by participation of the TNG LTA group in the Science WG of the AVO and in the Interoperability WG of the OPTICON project. The latter is a more technical group where standards are defined, discussed and agreed upon, and where the major data providers and the virtual observatories projects (AVO, NVO/USA, AstroGRID/UK) are represented.

5. Grid mechanism for accessing and processing archived data

Without considering the Grid paradigm, the most common mechanism to access archives, to retrieve data stored therein and to process the extracted data is based on a strict client-server scheme. As an example, one of the tools of the TNG LTA pilot project is an efficient, preliminary on-the-fly calibration of astronomical data stored within the LTA through a parallel Beowulf computing system. The facility is directly accessible through the Web interface of the LTA. By means of this interface users can activate a simple pipeline that allows the creation of pre-calibrated images and their subsequent download to the client computer where the user resides.

Within the Grid philosophy, the system created for the LTA prototype project is planned to become, in the OATs Grid node, the Storage Element (LTA archive and database) and Computing Element (Beowulf system). The LTA on-the-fly calibration will become a special service of the OATs Grid node, and any Grid user will be able to take advantage of it. More advanced user needs can be easily implemented by batch processes that can be configured through the User Interface to run on any grid node, accordingly to the runtime resources availability. This mechanism is shown in Figure 2. As a particular case, in the OATs Grid node a user can find the both the LTA archive + a Beowulf system so the system's performance can be optimized.

6. National and International status

A project called "Enabling platforms for high-performance computational Grids oriented towards scalable virtual organizations", approved and funded by the Italian Fund for Basic Research (FIRB), provides a technical and organizational framework so as to allow various projects to operate as virtual organizations. Within this project, three demonstrators for the astronomical community are included: access to the TNG pilot archive and GSC-II catalog, VST processing, monitoring of remote observations. From this basic kernel, core of an Italian Data Grid for Research in Astrophysics, an extension to new nodes providing data and applications is foreseen:

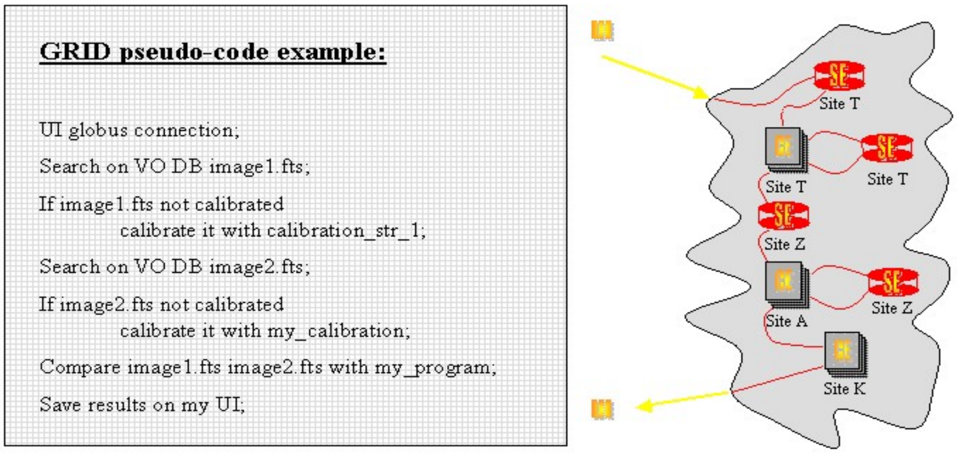

Fig. 2. An example of Grid pseudo-code and its mapping within the nodes of the Grid. It is to be noted that the user does not need to know which is the path followed at system level inside the Grid.

- access to data from high-energy missions and their reduction (ASDC and DIANA);
- integration of other applications (data processing, visualization, data mining and scientific) and extension in the number of nodes; a proposal for funding for this purpose has been submitted.

The project funded by FIRB has generated a national cooperation named IG-BIGEST (Italian Grid for Business Industry Government E-Science and Technology); the participation in this structure of the astrophysical community is guaranteed by INAF. IG-BIGEST participates in an EU FP6 project for the enabling of a pan-European grid for research (EGEE); the astrophysical community participates in its Task Force no.6 and in a dedicated Specific Support Action linked to EGEE. INAF participates in the EU FP6 Euro-VO project as well.

Interoperability among all VO data providers and repositories will be achieved by co-ordination within international working groups.

Acknowledgements. INAF participation in all EU activities is coordinated through its Third Department (Infrastructure and National Facilities); its Director, Giampaolo Vettolani, is warmly acknowledged for his suggestions and guidance. We also wish to thank Piero Benvenuti, Françoise Genova and Mirco Mazzuccato for their openness in discussing the needs of the Italian astrophysics community in the framework of European projects.

References

Benvenuti P., 2002 in: Organizations and strategies in astronomy III, Andr Heck (ed.), Astrophysics and Space Science Library, Vol. 280,107

Butler D., 2002 Nature, Volume 417, Issue 6891,777

Smareglia R., Pasian F.,. Zacchei A, Caproni A., Vuerli C., Longo G., Becciani U., Gheller C., 2002 in: Virtual Observatories, SPIE proceedings 4846,158

Pasian F., Smareglia R., Benacchio L., 2003 in: Toward An International Virtual Observatory, P.Quinn ed., ESO Astrophysics Symposia, in press

Lama N,, Vuerli C., Pasian F., 2003 in these Proceedings