Mem. S.A.It. Vol. 80, 477

Memorie della

The astrophysical cluster in EGEE

C. Vuerli[1,2], G. Taffoni[1,2], 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;taffoni;pasian]@oats.inaf.it

Abstract. The Astronomical and Astrophysical Cluster in EGEE aims at establishing and consolidating a well motivated astronomical community that makes use of the Grid technology. In this paper the status of the cluster is presented, focusing the attention on the cluster composition (participants) and planned activity.

Key words. Grid: applications – Grid: tools and services – Grid: EGEE project and infrastructure – Grid: EGEE disciplinary clusters – Grid: EGI (European Grid Infrastructure)

1. Introduction

The astrophysical cluster in EGEE was created in January 2007. In this paper we treat some aspects of the AA cluster. We put particular emphasis on the goals that we intend to achieve through the cluster. We illustrate therefore the current status by referring in particular the current composition in terms of participating people/institutes, the kind of clients that we expect to serve, the subtasks and activities that it is possible to identify yet at this stage, the metrics adopted to gauge the progress status, the applied procedures/policies, the communication plan, set up tools and organized meetings to support the community in terms of training and dissemination, some notes about the forthcoming activities and finally major issues that could raise during the two-years of the EGEE-III project lifetime and possible actions to mitigate the associated risks.

Send offprint requests to: C. Vuerli

2. Goals

The main goals that we want to pursue through the creation of the astrophysical cluster in EGEE can be listed as follows:

- Establish and consolidate a well motivated astronomical community that make use of the Grid technology. This objective can be achieved by informing astronomers about the potentiality that the Grid could offer for their everyday work by:
  - Offering a rich variety of hardware/software resources
  - Opening new opportunities to foster and strengthen scientific collaborations
- Demonstrate that the Grid is more and more helpful when different Grid infrastructures interoperate and different VOs support each other

Very often astronomers don't see any reason to leave well known technologies like the parallelization of their code on mainframes and local clusters and familiar software tools like MPI to approach a new technology like the

Grid that sometimes appears hard to use and not well consolidated yet. The key point and the main message that we should transmit to users is that the Grid could not and would not indiscriminately replace at all those technologies mentioned before. Rather its adoption should be evaluated carefully depending on the specific context in which the scientist operates from time to time. In any case the Grid could really represent an added value and be of great utility when different teams of researchers, remotely located cooperate in achieving a common objective usually in the framework of a big project. Scenarios like that envisaged above however are nowadays more and more frequent as scientific and technological objectives are ever more ambitious and require large collaborations. The astrophysical cluster therefore appears to be the right answer to this challenge in the framework of the EGEE project today and of the EGI (European Grid Infrastructure) in the near future.

3. Cluster composition

A number of astronomical Institutes of different European Countries are now part of the cluster. The list of the partner Institutes that compose the AA cluster is shown below. Partners are subdivided on the basis of the EGEE Regional Federations they belong to:

- Italy
  - INAF
- SWE (South West Europe)
  - IFCA
  - IFAE/PIC
  - ESA/ESAC
- Benelux
  - RUG
- CE (Central Europe)
  - SAS
  - UIBK
- SEE (South East Europe)
  - IPB
- France
  - OBSPM
  - OBSS/CDS
  - OBSL
  - OBSG
- DECH (Germay and Switzerland)
  - AIP
  - ARI
- UKI (United Kingdom and Ireland)
  - UCAM
  - ROE

Acronyms of the various Institutes are reported here: INAF (Istituto Nazionale di Astrofisica, IT), IFCA (Instituto de Astrofisica de Cantabria, ES), IFAE/PIC (Institut de Fisica d'Altes Energies/Port d'Informació Cientifica, ES), ESA/ESAC (European Space Agency/European Space Astronomy Centre, ES), RUG (Royal University of Groningen, NL), SAS (Slovak Academy of Sciences, SK), UIBK (University of Innsbruck, AT), IPB (Institute of Physics Belgrade, YU), OBSP (Observatory of Paris Meudon, FR), OBSS/CDS (Observatory of Strasbourg/Centre de Donneés Astronomiques, FR), OBSL (Observatory of Lyon, FR), OBSG (Observatory of Grenoble, FR), AIP (Astrophysical Institute Potsdam, DE), ARI (Astronomisches Rechen-Institut, DE), UCAM (University of Cambridge, UK), ROE (Royal Observatory Edinburgh, UK).

The EGEE-III, last edition of the EGEE project before the advent of EGI (European Grid Infrastructure), has been funded by the EU through the 7^{th} framework programme; a fraction of these funds have been allocated to support activities within disiplinary clusters. Unfortunately the pot of money reserved to AA cluster is not sufficient to finance all planned activities; consequently some of these activities could be cancelled or reduced. An unpleasant effect of the cut of funds is that some partner Institutions were left unfunded at all; despite this, many groups wish to continue their activity for the cluster, given its scientific and technological relevance.

The AA cluster is led and coordinated by INAF. The first author of this paper is the coordinator of the cluster whereas Dr. Fabio Pasian is the deputy coordinator.

4. Services

A rich suite of services is of crucial importance to make the Grid attractive for as-

tronomers. Astrphysical applications, in fact, are extremely challenging from the point of view of requested resources; these requirements can't be met by traditional Grids set up so far as they concentrate on computational aspects only. The concept of Grid as perceived by other disciplines however is something much more wide that includes non conventional resources and advanced services.

Services requested by AA applications can be of different nature. A list of the most important of them will be given below in this document.

It is worth to highlight however that as a general policy, the AA cluster will exploit tools and services built by the EGEE project teams and made available to all communities of users. This has the big advantage of a better coordination of the various disciplinary clusters and this is a key factor for the success of the project. As a secondary but not negligible effect, the waste of precious time and human resources is in this way avoided. In the following sections the main services requested by the astrophysical applications are briefly illustrated.

4.1. Documentation

A repository to collect all relevant documentation for the AA cluster will be created. Information that will be inserted in the repository includes:

- References to documentation repositories built by third party within and outside the EGEE project.
- General documentation for beginners who approach the Grid for the first time.
- General documentation illustrating techniques and best practices to take into account when porting applications in Grid.
- Specific documentation for users of the astrophysical cluster dealing with peculiar aspects of astrophysical applications that should help during the gridification process.

4.2. Hardware and software resources

As previously said, an adequate suite of hardware and software resources is very important to attract new users. To find the Grid attractive users should realize that they may count on a pool of new resources when joining a specificy community cluster and the related VOs, otherwise not available and able to make possible successful runs of their applications. The suite of resources include:

- Hardware: individuals and research groups will be encouraged to share their own resources within the AA cluster (i.e. to share them within the whole set or within a subset of the related VOs); as a further step these resources should be made available to other EGEE communities, so that astronomers can rely in return on resources coming from other communities.
- Software: the software that we plan to make available may be classified as follows:
  - Suite of scripts and procedures helping people in the process of preparing their applications to be run in Grid (e.g. application configuration scripts for different classes of applications).
  - Portals: Portals usually help in overcoming the psychological barriers interposed between the Grid and the community of users.
  - Grid M/W extensions to make the Grid able to fully meet the requirements of astrophysical applications (e.g. Databases and Remote Instrumentation);
  - Selection of AA applications among those ported in Grid as good demonstrators for dissemination purposes.
- Special attention within the cluster should be given to some hot topics related to AA applications. For each of these topics an adequate set of documentation, standards, best practices, tools and services shall be collected. Hot topics include HPC, Grid and the Virtual Observatory, Databases and Remote Control.

4.3. Training and dissemination

Very often AA applications dramatically improve their performances when submitted to a gridification process; many astronomers unfortunately don't have the necessary skill to carry out the gridification of their applications. To tackle this issue that could potentially hamper the achievement of the cluster goals, training and dissemination activities for people who join the cluster is of great importance.

This is certainly one of the key aspects for the success of the whole EGEE-III project. In this context a good set of documentation certainly helps but training events are essential to speed up the knowledge dissemination process. For this reason:

- Users will be encouraged to attend training events organized within EGEE NA4.
- Whenever necessary specific AA training events will be organized; in doing that we can count on the support given by EGEE NA4.
- Results achieved within AA will be disseminated to the whole project. This aspect is very important. The dissemination process will be carried out at a first stage by means of documentation repositories.

The final objective of these initiatives is to make users aware of real advantages in using the Grid for their work.

4.4. Clients

Clients (human users) of the AA cluster who typically take advantage of the set up services can be classified as follows:

- Grid site maintainers. These users need to access documentation related to techniques and best practices to set up and maintain a Grid site. Training events are also very important as they allow maintainers to get the necessary skill to set up and manage a Grid site in a very short time.
- Users who access the Grid to run third party applications only. This category of users mainly access documentation related to available hardware and software resources so that they can evaluate if requirements of applications they want to run are met. They are also interested in training events instructing how to use the grid and which available facilities can be expolited to run applications in the best and most efficient way with particular reference to Grid portals.
- Application developers. For application developers training events are usually helpful. These events focus on techniques and best practices to carry out the gridification process in the best possible way; in doing that application demonstrators are firstly considered as good test beds for the gridification process, thereafter the gridication process is undertaken for specific applications proposed by trained people.

5. Subtasks

The detailed workplan for disciplinary clusters in EGEE-III has not been defined yet. Deatiled workplans will be scketched out after the starup of the project. The detailed workplan for AA cluster will be prepared with the contribution of all Institutes being part of the cluster. It will be therefore further discussed at the EGEE kick off meeting, thereafter it will be consolidated. It is worth to say however that the workplan is a living document and it will be revised and updated as things evolve within the cluster in terms of applications, tools and services, as well as in terms of participating institutes. Yet at this early stage however it is possible to identify a first kernel of subtasks that will be detailed at a later stage. These subtasks include:

- Documentation. Collection of useful documentation from different sources (especially from NA4) and possibly set up of a specific documentation repository for AA cluster.
- Hardware. It is extremely important to foster the sharing of new resources within our 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 and therefore the suite of functionalities offered by the Grid.
  - Census of gridified AA applications to select good AA demonstrators to be used for training and dissemination purposes
  - Selection of tools developed within EGEE and useful for AA applications
- Training and dissemination. For this subtask it is important to say that:
  - The EGEE project provides training services for all disciplinary clusters, both for Grid novices through introductory training events and for Grid application developers.
  - It is important to clearly identify the needs of the AA community.
  - AA specific training events (addressing specific astrophysical topics when approaching the Grid) are very important. They in turn can be organized with the support of the related EGEE teams.

6. Topic NA4 activities

In this section some topic activities (listed below) of EGEE NA4 are reported. The goal pursued by NA4 can be summarized as follows.

- Support for virtual organisations (VOs) and for their users. In this context three distinct teams provide:
  - "Virtual Organisation Support" aimed at easing the management of users within a VO as well as the wide and increasing number of VOs using EGEE.
  - "Application Porting Support" to help developers in effectively porting their applications to the Grid.
  - "Direct User Support" to help users in solving day-by-day problems when using the Grid.
- Creation of "Strategic Discipline Clusters" that play a central role in setting up services provided by the infrastructure.
- "Community building" aiming at minimizing the effort required to support Virtual Organisations and users that wish to exploit the EGEE infrastructure. The objective is to make disciplinary communities as self-reliant as possible.

7. Metrics

It is important to establish a number of parameters whose values allow to gauge the progress of activities within the cluster over the time. A first set of parameters that could be useful for this purpose is reported below. This preliminary set will be probably revised at a later stage.

- Number of Institutes, groups and individuals who join the cluster
- Number of users who registered the generic AA VO and other related VOs
- Number of gridified applications
- Number of submitted jobs.
- Number of jobs whose runs terminated successfully
- Number of failed jobs
- Number of shared resources
- Number of attended and/or organized training events

8. Procedures and policies

The purpose of procedures and policies is to control access to the various resources and services by the community of users who joined the cluster. Access control concerns both a mapping between users and resources/services and how each resource/service can be used by allowed users. Procedures and policies could be set up for the following resources and services:

- Access to hardware and software resources.
- Adoption of already developed new software in terms of scripts or extensions of the Grid M/W.

- Proposals to choose a specific gridified application as pilot application (demonstrator).
- Ask/offer support for what concerns training events.
- Propose new software tools that could be suitable to achieve the cluster goals.

No particular restrictions however are envisaged at present in giving access to these resources and services. Proper forms will be eventually prepared and made available on the cluster web site if some access restriction will be applied in the future.

9. Communication plan

The astrophysical cluster, like any other disciplinary cluster in EGEE, is spread over different European countries and many institutes participate to it. Efficient communication mechanisms therefore is a necessary precondition for the success of planned activities. We plan to set up and use the following tools to allow people to communicate within the cluster:

- Web based communication mechanisms. The idea is to set up a web site dedicated to the astrophysical cluster. The web site will be based on a collaborative tool (almost certainly TWiki but also Joomla could be a good candidate) so that all cluster members can contribute to refresh and maintain the web pages.
- Mailing lists. In the early phase of the EGEE project a single mailing list should be enough. More mailing lists (e.g. one for each subtask) could be set up afterwards if this enforces the collaboration between scientific groups.
- Internet based audio/video conferences play an important role within the cluster. Funds allocated to disciplinary clusters, in fact, are entirely invested to acquire manpower. This could be source of trouble for people who need to afford travel expenses to attend face-to-face meetings. This unpleasant situation should be mitigated by making use of remote audio/video conferences.
- A ticketing system tool (GGUS or others).
- SVN or CVS as a collaborative tool for concurrent software development and management. A tool for the generation of software packages will also be chosen.
- A mailing list server tool.

Contacts with other tasks of NA4 are very important for a a number of reasons the most important being organization and exploitation of support activities, tight contacts with other strategic clusters and coordination with the rest of the EGEE project.

It is planned to set up software and documentation repositories. Alternatively, those provided by the project will be used.

10. Future plan

The EGEE-III project starts on May 1^{st} 2008. A kick off meeting focused on the definition of the workplan for each specific cluster is expected. The workplan of the AA cluster will be prepared with the contribution of all members. The cluster workplan will be prepared to define in detail all activities that will be undertaken within the cluster.

For each identified main activity a subtask will be created. Each subtask shall have associated a unique ID, the activation and termination dates (in months starting from T_0, where T_0 is the starting date of the project), the institutes and people that will contribute to the subtask with the role for each of them, and finally a description of the goals and benefits of each subtask.

The first draft of the subtask will be initially prepared through remote interactions among all involved Institutes and people; the first draft of the workplan therefore will be discussed at the kick off meeting and a first consolidated version of it will be released.

The workplan is the main document of the cluster; it is a living document as it will be continuously kept up to date to reflect the evolution of the cluster in terms of contributors, applications, tools and services.

Some milestone face-to-face cluster meetings will be organized. To make possible a massive participation of the cluster members, mailstone meetings will be organized jointly

with the most important EGEE events, namely the yearly EGEE User Forums and EGEE conferences. All cluster members however will keep constantly in touch through the collaboration tools (see section 9).

11. Issues and risks

Certainly one of the most important issues for what concerns the AA cluster is the shortage of allocated funds. This aspects could make difficult to achieve some goals of the cluster. In particular: a) some problems could arise in hiring the necessary man power to develop new software services and to port applications in Grid; b) people could have problems in procuring funds to attend the cluster face-to-face meetings. In order to mitigate risks coming from these issues tools and services made available by the EGEE project will be used whenever possible. The lack of funds for mobility shall be partially mitigated by making massive use of internet based audio/video conferences like SkyPE, EVO and GoToMeeting.

12. Conclusions

The creation of the astrophysical cluster in EGEE is a big opportunity for astronomers and astrophysicists to exploit the rich set of resources made available for communities that participate to the project. AA users have the chance to really experiment that the Grid can be the right answer for certain classes of applications particularly challenging in terms of requested resources, tools and advanced services.

EGEE with its Grid infrastructure can really foster scientific and technological collaborations in wide and geographycally spread projects. To make the EGEE-III project as fruitful as possible, activities of the AA cluster must be carefully detailed in the workplan. The workplan must clearly indicate who will contribute to what (in terms of subtasks) and the effort (in terms of PM) that each member can guarantee to the various subtasks.

A suite of proper tools and services is of vital importance to make effective the collaboration and the flow of information and data within the cluster.

If successful, EGEE-III is the project through which the European Astrophysical Community can establish a strong and solid presence in EGI, the forthcoming European Grid Infrastructure.

Acknowledgements. A special acknowledgement is due to Charles Loomis, Frank Harris and Vangelis Floros of the EGEE NA4 Steering Committee who made possible the construction of the astrophysical cluster.

We want to acknowledge all EGEE Regional Federations that supported the man-power requests for the astrophysical cluster.

Finally we wish to thank all colleagues (single astronomers and astronomical research groups) who accepted to be part of the AA cluster and operate to make this experiment successful.

References

Jones, R. 2007, Enabling Grids for E-sciencE, 1, 228.

Vuerli, C. 2007, Proposal for the Creation and Development of the Astronomy and Astrophysics Cluster in EGEE-III, v. 3.1, 1, 36.