Generic Applications Questionnaire

Introduction of Grid technology for the MAGIC telescope
First step towards a Europe wide Astroparticle Grid ASTROPA

Harald Kornmayer, Forschungszentrum Karlsruhe
for the MAGIC collaboration
harald.kornmayer@iwr.fzk.de

General Comment:
Most of the information about the application we want to bring to EGEE can be found in the separate document which I send to you too. We are already working together and trying to realize the first application for the MonteCarlo Production for one dedicated experiment. Therefore I addressed my answer mainly to this application. This application will be the prototype for the community to show them the benefit of Grid technology. For this application I see the biggest potential for synergies between the different astroparticle experiments. I didn't answer all question concerning the analysis for the data from MAGIC. This will be the second step of the project. After a successful start and a move from the MonteCarlo prototype to a MonteCarlo production system.

Section A: Describe your community

1) Is your community scientific or industrial? The astroparticle physics community, where the proposed project addresses, to is purely scientific. The aim of a European wide ASTROPA Grid offers the community a tool for collaborative work on compute intensive and data intensive applications. The ASTROPA Grid would ease the analysis of data from different experiments from satellites via ground based cerenkov telescopes up to big air shower experiments like AUGER or big neutrino telescope like Antares or ICECUBE. The grid would support the current leadership of Europe in the air cerenkov technique.

2) Describe concisely the "added value" for your community to run your application(s) at an European-wide scale specifying in particular what could be done on the grid infrastructure and could not be done at a local scale (max 500 words).
The simulation of the cosmic particles is a big challenge. For example the production of the hadronic background for the MAGIC telescope need – according to the recommendation of the INFN overview board for the INFN MAGIC group in Padua and according to the current production speed – needs 11400 2 GHz Pentium CPUs for one year. The same situation occurs for a satellite like GLAST where the production of MonteCarlo events needs more

the 7000 CPU days (2 GHz Pentium). The data challenge 2 of GLAST wants the equivalent amount of MonteCarlo data as for one month of observation. This kind of simulations is necessary to understand the dominant background of hadronic particles. Without the computing power of the Grid this goal is unreachable for local research groups. Also the data management of this amount of data is a challenge that needs the support of a distributed storage management system available at the Grid. The biggest value might be the possible joint collaboration between different experiments to analyse on sky source within different wavelengths.

3) In how many countries (N) in Europe is your community spread out ?
   a. N<5
   b. 5<N<15
   c. N>15

   The probable might be between 4 and 15 countries in Europe, also from the new member states.

4) How many people (N) in your community will be using a grid infrastructure in the next year ?
   - N<10
   - 10<N<100
   - N>100

   If the proposed infrastructure is available, the number of users will be more than 10 and less than 100.

5) How many people (N) in your community will be using a grid infrastructure in the next two years ?
   - N<10
   - 10<N<100
   - N>100

   The number will be increasing, depending on the success of Grid at all, but it will not exceed 100 to much.

6) From how many sites (N) in Europe will people belonging to your community connect to the grid infrastructure in the next year ?
   a. N<10
   b. 10<N<50
   c. N>50

   The proposal foresees the integration of around 6-8 sites in the MAGIC Grid, including in the national Grid sites in Italy, Spain and Germany.

7) From how many sites (N) in Europe will people belonging to your community connect to the grid infrastructure in the next two years ?
   a. N<10

b. 10<N<50

c. N>50

Due to the number of different experiments the number will increase and be between 10 to 50. The contribution will be different experiments and different sites.

8) Describe concisely (max 500 words) the security requirements of your community.

The security of all distributed system is an important topic. But the astroparticle physics community is purely scientific and therefore the security problems will not be as great as in field like medicine, pharmacy or genomic applications. This offers the use of already existing security mechanisms like the Globus Security infrastructure, etc. without any longer time lag due to not available stricter Security and Privacy mechanisms.

Section B: Describe your application no. 1

9) Is your application no. 1 scientific or industrial ?

Scientific.

10) Describe concisely your application no. 1 (max 500 words) from the point of view of its goals and algorithms.

The first application is the simulation of high energetic particles penetrating the earth atmosphere and the subsequent so called air shower. This shower includes the simulation of particles and cerenkov photons. The simulation is based on the standard simulation program called CORSIKA used by all ground based air shower experiments in the world. This program uses standard simulation algorithms from the high energetic particle physics community. After the air shower simulation the workflow continues with the simulation of the experiment's detectors. This simulation is based on ray tracking algorithms and standard high energy physics programs. Because of the big number of particles produced for the high energetic primary particle, the simulation produces a lot of intermediate data.

11)Is your application no. 1:

a. CPU intensive ?

b. data intensive ?

c. both ?

The application is mainly CPU intensive. But the management of the (intermediate) data will be a challenge too!

12) Is your application no. 1 mainly intended for:

- interactive use ?

- batch use ?
- both ?

This application is intended for batch use.

13) Quantitatively evaluate your application no. 1 in terms of:

a. CPU power (SpecInt2000/SpecFp2000 per second per job);
b. Memory consumption per job (Megabytes);
c. Disk storage needs per job (Terabytes);
d. Tape storage needs per job (Terabytes);
e. Number of jobs per user per year;
f. Number of users per year;
g. Network bandwidth requirements (Megabit/sec).

Here I want to start with a general comment. Some of the questions are not easy to understand. The number of disk storage needed depends on the number of CPUs one gets for production. The proposed application is easily to work in parallel processing. So a big job can be distributed on smaller subjobs. So the CPU power per job is – at least for me – not an interesting number. But let me try to answer your question nevertheless by addressing all of the above mentioned terms:

a.) The needs for the simulation of hadronic cosmic ray shower to study the background of the MAGIC telescope or the GLAST satellite is very big (see above, 10000 CPU years or 7000 CPU days on a 2GHz Pentium CPU). I don't have precise number on the specs currently.
b.) Memory consumption < 500 MB
c.) Temporarily disk space: 20 TB
d.) Persistent storage space: increase of 1.7 TB per CPU per year.
e.) The number of jobs depends on the number of CPUs available. The aim is to allocate up to 200 dedicated CPUs in the first year and the use of idle CPUs of the Grid to get an efficient use of the money spent by the tax payers.-

14) Quantitatively evaluate which percentage of the data of your application no. 1 needs to be replicated in more than one site and the average number of copies per elementary replicated data set (e.g., file).

The design of the system tries to store the data in three data centres and to allocate the job close to the data (bring job to data!). Therefore not more than 10% of MonteCarlo data needs to be replicated.

15) Does your application no. 1 have a graphic or a command-line user interface ?

Currently the application is started with a command line user interface, but efforts are ongoing to integrate them in a GUI by using portal techniques or the Migrating Desktop from CrossGrid.

16) Can your application no. 1 be accessed/steered from within a web browser ?

Not yet, but the access via web server is already addressed.

17) Is your application no. 1 already interfaced to any grid middleware ?

First runs are done on the CrossGrid testbed. There is work ongoing to run int on the CNAF grid within 2 weeks for first tests.

18) If yes to question 17, to which middleware and in the context of which project?

The used middleware is based on LCG-2 in the context of CrossGrid and EDG.

19) If yes to question 17, can you cite some references (less than 10) to related work ?

www.crossgrid.org,
http://eu-datagrid.web.cern.ch/eu-datagrid/

20) Did people of your community already attend any grid demos/tutorials ?

The project is based around some national grid centres with a lot of grid experience. The system will be set up by them and the people from the community will be teach by them – hopefully with support form NA3 and NA2.
FZK is joining the NA3 activities and will present the EGEE idea to the MAGIC collaboration.

21) If yes to question 20, how many people per event ?

Enough experience in the Grid centers CNAF, PIC and FZK. But at least one joined already the “EGEE induction course” at CERN and “Gilda tutorial” in Catania.

21) If yes to question 20, how many events of each kind (demos, tutorials) ?

One of each of them – demo and tutorial.

22) Does your application no. 1 need third party commercial software to run ?

NO!

23) If yes, which one(s) ?

24) Describe the license under which your application no. 1 can be distributed on the grid infrastructure (max 200 words).

Up to know the usage of the software is free for scientific applications. But the license issue was not discussed in details. The project will follow a open source lisence. But due to the experience in some EU projects we recommend to ask for a EU Grid general license. We are yet not convinced that the EDG license is fine for all cases. But it might be the prototype for the license.

25) Describe concisely (max 500 words) the security requirements of your application no. 1.

See above.

26) If yes to questions 17 and 20, describe concisely (max 500 words) what problems you foresee to successfully port your application no. 1 on a distributed computing environment like an European grid infrastructure.

The problems for this application is the set up of a VO and the allocation of resources for this VO. Technically we don't expect to much changes on the current application programs. Some work will be necessary to bring the GUI to a productive level by using portal , but no principle problems will arise. The main challenge – as in all productive distributed systems – will be the administrative coordination within the project and without the providers of infrastructure like EGEE and the national Grid centres.

Section C: Describe your application no. 2

As already mentioned in the general comment to the questionnaire at the beginning of the document the second application might be a analysis system for the MAGIC telescope. This application would be more data intensive and the first version of the system will be based on standard air cerenkov telescope technologies. But the Grid might offer also the possibility to introduce new analysis methods like neuron nets, advanced fitting algorithms, etc to strengthen the leading position in air cerenkov telescope technique in Europe.