



Contribution ID: 158

Type: **Poster**

## **Grid User made easy: Creating generic script for specific problems**

*Monday, 11 April 2011 09:00 (8 hours)*

### **Overview**

The Grid Group of the Institute of Astrophysics of Andalucía (IAA) work is mainly focused in software simulations, data analysis observatories or image filtering. Problem: We have noticed that Grid users think that this infrastructure is very difficult to use and it is not 100% safe because of its heterogeneity. Users must know many commands lines and use them all the time. They will use them if they want to submit jobs, test the job status, handle data or monitor the Grid infrastructure. A great effort is necessary in order to learn and to implement each command. Solution: To solve these problems we have created a group of tools that synthesize the main commands used in the gLite middleware. We have analyzed the use cases, designing, implementing and testing a group of shell scripting language to make the sending, monitoring and data collecting of large batches of data from the Grid infrastructure easier. These powerful general scripts can be modified for special cases.

### **Impact**

The impact of this work focuses on all areas where these scripts have been modified for. The Research Departments are Extragalactic Astronomy, Stellar Physics, Radioastronomy and Galactic Structure and Solar System. This work has had a major impact in the areas of astrophysics where they have been tested. In general terms, the astrophysical applications launched an average of 1000 jobs per run. So the estimated cost of improvements is as follows: Time: Using Grid and parallel jobs, the range of improvement over time is between 40% and 80% depending on the degree of parallelization. This improvement is due to the use of Grid and not to the use of the scripts. Number of commands: This improvement is due to the use of the implemented script. For example: A group of one hundred standard jobs in the case the user don't use the scripts. Then he should used 1000 command lines to prepare the structure, 100 command lines to submit jobs to the Grid infrastructure, 200 commands lines to check the job and relaunch the wrong jobs, 500 command lines to collect the results and put them in the correct output structure and 100 command and 100 command lines to remove groups of works. To sum, the user have to used about 1900 command lines without scripts. In the other case, using scripts the total command lines using scripts are 5. Effectiveness is 95% earned in addition to the usability of the Grid Infrastructure or large groups of jobs. Cron: The cron user should be continually reviewing the status of job. So, the time spent in this task is significantly improved using this tool.

### **Description of the work**

The work we have developed throughout the last two years can be summarized in the creation of a group of general tools for data managing, jobs sending, status monitoring and results collecting on the Grid infrastructure. As extra work, our group has modified these tools to adjust them to specific astrophysics cases, adapting these scripts to port all the astrophysical applications to the Grid Infrastructure. This effort was rewarded

when astrophysical scientists improved their expectations to use grid. So, there are two examples of solved problems using our scripts. The first one is HEALPIX (Hierarchical Equal Area isoLatitude Pixelization of a sphere). The second one is DDSCAT (The Discrete Dipole Approximation for Scattering and Absorption of Light by Irregular Particles). The general scripts and their descriptions are: Prepare: Creates an input data structure containing all necessary files for sending a group of jobs to the Grid platform. Run: Sends Grid infrastructure jobs created by the script "prepare". Verify: Checks the status of the submitted jobs and if they are in a wrong state (Cancelled or Aborted) resend them, thus the user does not have to worry about corrupted work. Collect: Collects the results and sorts them into the created structure when the state is "Done". Cancel: Cancels the group of jobs created by "prepare" script. This script is not a mandatory but it is very necessary. Cron: It is necessary to use the cron script to use the former ones more efficiently. So, the end user should not be worry about the state of the work nor the location of the final files. This cron file is created by the script "prepare", it is launched by the script "Run" and it runs the script "verify" very often. If the job is failed, the job will be launched again. If the job is done, the script "collect" will be launched and it will put the results in their proper place.

## URL

<http://grid.iaa.csic.es>

## Conclusions

The usefulness of some scripts that enhance and help scientists in their work has become evident. The scripts are not only used as a tool for massive use of the Grid, it is also used as a tool for monitoring and fixing bugs.

The usefulness of some scripts that enhance and help scientists in their work has become evident. The scripts are not only used as a tool for massive use of the Grid, it is also used as a tool for monitoring and fixing bugs. Using these scripts, users improve the 95% of time when the jobs are sent to the Grid infrastructure. Users do not have to train or learn the necessary orders to use the Grid. These scripts are ideal to launch working groups with a lot of computing requirements. Scientists using the scripts developed by our group will notice that the fact of sending jobs to the Grid infrastructure has become much easier as they do not need to know the basic Grid commands.

**Primary author:** RODON ORTIZ, José Ramón (CSIC)

**Session Classification:** Posters

**Track Classification:** Poster