XMM-Newton Data Analysis with SAS Software over GRID Technology

M.T. Ceballos, I. Campos, P. Orviz, D. Tapiador, R. Álvarez, A. Ibarra, C. Gabriel, and J.R. Rodón

Abstract The processing of data obtained by the XMM-Newton observatory of the European Space Agency (ESA) is done using the SAS (Software Analysis System) tools provided by the ESA XMM Science Operations Centre (SOC). In order to be operative, these tools must be downloaded from the official SOC web pages and then installed (eventually upgraded) and run locally in the computer used by the data analyser. This process can be in some cases cumbersome for some users and local resources. In this presentation, we summarise the initiative developed from the SOC at the European Space Astronomy Centre (ESAC) in collaboration with the Instituto de Fsica de Cantabria (IFCA CSIC-UC), to run these tools in a GRID environment, and thus taking advantage of the resources distribution and the possibility of remote use of the analysis tools.

M.T. Ceballos
IFCA(CSIC-UC), Avda Los Castros, Santander, e-mail: ceballos@ifca.unican.es

I. Campos
IFCA(CSIC-UC), Avda Los Castros, Santander, e-mail: iscampos@ifca.unican.es

P. Orviz
IFCA(CSIC-UC), Avda Los Castros, Santander, e-mail: orviz@ifca.unican.es

D. Tapiador
ESAC(ESA), Madrid, e-mail: Daniel.Tapiador@sciops.esa.int

R. Álvarez
ESAC(ESA), Madrid, e-mail: Ruben.Alvarez@sciops.esa.int

A. Ibarra
ESAC(ESA), Madrid, e-mail: Aitor.Ibarra@sciops.esa.int

C. Gabriel
ESAC(ESA), Madrid, e-mail: Carlos.Gabriel@sciops.esa.int

J.R. Rodón
IFCA(CSIC-UC), Avda Los Castros, Santander, e-mail: rodon@ifca.unican.es

1 The XMM SAS tools

The Software Analysis System (SAS) [1] is a set of analysis tools developed by XMM-Newton Science Operations Centre (SOC) at ESAC (European Space Astronomy Centre) in collaboration with the Survey Science Centre (SSC) consortium, to process the data obtained by the XMM-Newton observatory.

The standard processing with a chain of these software tools generates standard calibrated products such us event lists, the spectra and light curves for the brightest sources, images, exposure and background maps in different energy bands, etc. These products are delivered with the raw data to the observer and after one year of proprietary rights, incorporated to the public archive (XMM Science Archive [2]).

The observers or the users of the data archive can then generate new scientific products or even re-generate the standard ones (starting from the raw data) to take advantage, for example, of the most recent calibration information.

1.1 SAS processing

Currently, the XMM data must be processed by the scientist using his local computing resources. The SAS software must be downloaded, installed and run locally thus facing some (occasionally non-trivial) problems:

- Auto-installation and maintenance of the SAS and additional software
- Download and updates of the calibration database
- Processing resting in the local resources: the data analysis can be computationally demanding if the number of observations to be processed is large or the observations are long in time

2 Distributed Resources: the SAS GRID initiative

The alternative to run the analysis software using the local resources is taking advantage of the GRID infrastructure, that is using the many computational resources distributed along the research centres involved in any of the E-Science projects. This initiative has been started from the SOC at ESAC, in collaboration with the Instituto de Física de Cantabria (CSIC-UC), through the EGEE (Enabling Grids for E-Science) project where both institutions are members.

The EGEE project [3] is funded by the European Commission and it is defined as "the largest multi-disciplinary grid infrastructure in the world, which brings together more than 140 institutions to produce a reliable and scalable computing resource available to the European and global research community".

Among the many Virtual Organisations (VO) in which the EGEE project is structured, the SAS GRID work has been developed inside the "Planck" VO, since both

collaborating institutions (ESAC and IFCA) are current members. The possibility of creating a new VO as a frame for this work is still under study.

2.1 SAS GRID collaboration

The collaboration between ESAC and IFCA involves all the aspects of software installation as well as the GRID authorisation mechanisms and users support. In particular:

- SAS (and external) software installation
- Installation and automation of the calibration database
- GRID Node identification as "SAS" members using "tag" concept so that the resources discoverer can be aware of the SAS software availability
- User support based on the elaboration of (electronic) manuals, training sessions and development of meta schedulers that could make easy the user approach to this (new) environment

The technical data of the computational resources dedicated to the SAS GRID project in both centres can be summarised as follows:

- IFCA
  - 350 CPU x Intel Xeon, 3/3.2 GHz
  - 2-4 GB RAM
  - 40 TB storage
  - Middleware: EGEE-gLite 3.1
  - OS: Scientific Linux 4
  - Computing Element: LCG (globus)
- ESAC
  - 15 x Dell blades (7 with 2-Dual Core CPU, 3.2 GHz, 6GB RAM and 8 with 2-Quad Core CPU, 2GHz, 16 GB RAM
  - 140 GB storage (RAID 1)
  - Middleware: EGEE-gLite 3.1
  - OS: Scientific Linux 4
  - Computing Element: LCG (globus)
  - Meta scheduler: GridWay

2.2 User approach

Currently, a user (an astronomer) can process his data remotely using distributed resources if only two conditions are fulfilled: he is member of the Planck virtual organisation (permissions, certificates) and he has an account in a user interface

at any of the centres involved (IFCA, ESAC). The advantages of this new way of processing are obvious: the user has access to a lot of processors without any need of local installation of software, calibration and maintenance.

3 Future

The near future of this project resides in the final publication of the RISA (Remote Interface for Science Analysis [4]) project. RISA is a client/server application that lets the user run the SAS processes remotely on top of a GRID architecture and through a GUI interface (see 1). This system facilitates the job submission and is able to connect directly with the XMM archive to retrieve that data to be processed.

References

1. http://xmm.esac.esa.int/sas/8.0.0/
2. http://xmm.esac.esa.int/xsa/
3. http://www.eu-egee.org/
4. Arvisetc C. et al.: Astronomical Data Analysis Software and Systems (ADASS) XVII, 394 $http://www.aspbooks.org/a/volumes/article\_details/?paper\_id = 29005$

Fig. 1 RISA design.