Highlights of Spanish Astrophysics VI, Proceedings of the IX Scientific Meeting of the Spanish Astronomical Society held on September 13 - 17, 2010, in Madrid, Spain. M. R. Zapatero Osorio et al. (eds.)

Towards the first open repository in gamma-ray Astronomy

Jelena Aleksić[1], Ignasi Reichardt[1], and Javier Rico[1,2] (on behalf of the MAGIC collaboration)

[1] Institut de Fisica d'Altes Energies, Edifici Cn, Campus UAB, E-08193 Bellaterra, Spain
[2] ICREA, E-08010 Barecelona, Spain

Abstract

The MAGIC Telescopes produce around 300 TBytes of raw data per year that are pre-processed on-site at the Observatory *Roque de los Muchachos*, La Palma, and then transferred to the MAGIC Data Center at Barcelona, for storage, processing and internal data access by MAGIC users. MAGIC Data Center is supported by *Institut de Física d'Altes Energies* (IFAE), *Universidad Complutense de Madrid* (UCM), *Instituto de Astrofísica de Andalucía* (IAA) and *Universitat Autònoma de Barcelona* (UAB), and hosted by *Port d'Informació Científica* (PIC), Barcelona. We discuss here the possibilities that the MAGIC Data Center offers for the data handling, as well as the undergoing modifications aimed at further improvement of its performance. In addition, we discuss the accessibility of the MAGIC data to the wide scientific community. Aside from providing all its published data in a format widely accepted by the astronomical community (FITS), the MAGIC collaboration is currently in the process of adapting these data to the standards required by the International Virtual Observatory. This will put MAGIC published results on the international network of astronomical data and make them available to the whole scientific community.

1 Introduction

MAGIC (Major Atmospheric Gamma-ray Imaging Cherenkov) is a two-telescope ground based system that detects and studies VHE gamma-rays. Telescopes are located at the Observatory *Roque de Los Muchachos*, on Canary island La Palma, at an altitude of 2200 m. The first telescope, MAGIC-I, has been taking data since 2004, and the second one, MAGIC-II, is operational since 2009. Together, they produce hundreds of TBytes of data per year, and the computing power needed for processing them is measured in years of CPU time. Production, storage and processing of this large amount of data represents a great technical

challenge, and within the MAGIC collaboration it is dealt with through the MAGIC Data Center.

MAGIC Data Center was set up and is run by groups at IFAE, UAB (Barcelona), UCM (Madrid) and IAA (Granada). It is located at PIC, Barcelona, and is equipped with the storage resources and computing capabilities needed to process the data from the telescopes and to make them available to the MAGIC collaboration. Following the upgrade of the MAGIC-I telescope that is scheduled for the next, the volume of recorded data will be increased, and as a consequence, the Data Center will have to augment its capabilities, both through hardware and human resources. The Data Center also provides access to the MAGIC high-level public data, that are stored in a widely-used FITS format and are available to everyone. In addition, Data Center will also host a link with the Virtual Observatory (VO), making the MAGIC data first VHE data available through VO.

In the following we will describe the present status of the Data Center and the foreseen upgrades required to meet the needs of the MAGIC Telescopes. We will also discuss the plans to provide additional services to the collaboration members and to the scientific community.

2 MAGIC data: production and analysis

The MAGIC telescopes detect Cherenkov radiation (collected by the mirrors) by focusing it to the camera that is segmented into a number p of pixels (577 for M-I and 1039 for M-II) [1]. If a signal is strong enough to provoke a trigger, charge contained in every pixel is sampled s times (50 M-I, 80 M-II). Each sample is digitized with a 12 bit precision, and the resulting values stored in 2 byte fields. In addition, a fixed-size (4.5 kByte) header is written to describe the event's properties, amounting to p*s*2 B + 4.5 kB of data volume recorded by each telescope per event. If we also include the data acquisition rate (200-700 Hz) [4], and multiply it by the number of observational hours performed by the telescopes per year ($\sim$1500 h), we get that the amount of data recorded by MAGIC Telescopes is around 300 TBytes per year.

These data are saved as RAW files and, in order for analyzers to use them, they have to be processed further. For that we use MARS (MAGIC standard Analysis and Reconstruction Software) —a set of ROOT-based programs that form part of the analysis chain or are used for production of high-level results [2].

First step is calibration of the data: the pulses registered by FADCs are integrated and converted into the number of photoelectrons and arrival time. By doing so, instead of s numbers recorded for every pixel in RAW files, only 4 per pixel are saved in the calibrated ones (charge, arrival time and their uncertanties). This part of the analysis is the most CPU demanding.

After that, the calibrated files are taken as the input for the next step in the analysis chain: reduction. Here, the image cleaning and parametrization are perfomed, and as a result we get values describing the shower itself: its estimated energy, nature (how gamma-like is the shower) and direction to its origin. After the reduction, data can already be used for production of high-level analysis outputs (light curves, spectra, skymaps, ...).

Table 1: Data volume for the different phases of the MAGIC Telescopes analysis chain

Program	Compression[a]	Data Volume[b]
Data Acquisition (RAW)		300 TB/yr
Gzip	0.3	100 TB/yr
Calibration	0.034	10 TB/yr
Reduction	0.0026	1 TB/yr

[a]Compression factor w.r.t the RAW data volume
[b]After the MAGIC-I upgrade, this number will increase

RAW files, as well as outputs of calibration and reduction processes are stored permanently.

Estimations the data volume at the different stages of the analysis chain are listed in Table 1.

3 Data center

Currently, the MAGIC Data Center takes care of the following tasks:

- Data transfer from La Palma, via Internet and tapes.
- Data storage on tapes and disks at PIC.
- Data access to all data processing levels (raw, calibrated, reduced) all MAGIC collaboration members.
- Real-time, automatic analysis of the data, processed with the MAGIC standard software.
- Resources and tools for private analyses by any MAGIC collaborator, including access to data on tapes.
- Reanalysis of all stored data in the case of software updates or bug fixes.
- Software repository and bug tracker.
- Storage of the data quality control files.
- Hosting the MAGIC Public Data Base.

During data taking, the RAW data are stored on the disk on site, and later recorded to the tape. The tapes are then sent from La Palma to PIC via airmail, since the Internet connection between the island and Europe mainland still does not have sufficient bandwidth.

Also, the computing system on site performs the so-called "On-site" analysis [3], that processes the RAW files into calibrated and reduced ones. These files are transferred to PIC as well, but via internet, together with some log files generated by the subsystems of the telescopes.

Data at PIC are organized in a database and served to the Collaboration through a user interface machine and a web site. Until late 2008. these data were in an NFS file system, that is now replaced with a new GRID-based file system (dCACHE), that allows transparent access (in the sense of making no difference between the tapes and disk storage) to any level of RAW and processed data.

The reduction of the files that arrive via Internet is triggered by the scripts that run automatically every few minutes and notice when the transfers have finished successfully. When that happens, batch jobs are generated and submitted to the GRID with a specific configuration that ensures that they will run at PIC. This set of scripts can also be used to reprocess all the data stored at PIC in a case a bug in the software is found, or some change in the software is made that significantly improves the analysis.

The Data Center also provides a "Concurrent Version Server" (CVS) for the software development and the daily check, which generates a daily report on the data quality and the telescope stability.

3.1 Public database

The Data Center also hosts the MAGIC public database, where the published MAGIC results are stored in the widely-used FITS format. The database is accessible by everyone through a web page that allows for search and download of the FITS files. The files themselves contain high-level outputs of the analysis (like skymaps, spectra, phaseograms, theta2 plots, light curves...), but also some additional information describing the source, observations and analysis setup. We have introduced some special keywords to the FITS format to better describe the taking and processing of the data. At the database's web-page there is also a short manual describing all the new terms used in our FITS files.

In the past few months we have been working on making the MAGIC public data Virtual Observatory-compliant. In collaboration with people from LAEFF (Spanish VO), we are currently developing the necessary tools that will allow accepting and processing VO queries, and returning the search results to the user. So far the tools for spectra and light curve searches have been completed, and are in test phase. Given that the MAGIC public data will be the first VHE data available through the VO, we will still have to develop some new protocols to represent the results that are only characteristic for gamma-ray astronomy.

4 Future plans for the data center

In the near future we want to provide additional services to allow a more agile analysis by any MAGIC collaborator and also easier access of anyone to the published data. Our next steps include:

- Deployment of the new hardware for the data center (increasing the computing power by a factor 10, storage on disk by a factor 3 and tape access speed by a factor 4).
- Optimization of the reanalysis process (goal: reprocess 1 year of data in 5 days).
- Transformation of the data center into a computing center, where users can not only access the data, but also run their analyses over them.
- Providing user-friendly tools for data analysis.
- Continuous update of the FITS Public database.
- Developing tools to allow for image search according to VO's SIA protocol.
- Opening the MAGIC public database to the Virtual Observatory.
- Providing access to the data and analysis tools for external observers (still under design study).

Finally, given the future merging of the MAGIC with other imaging air Cherenkov telescope experiments into the CTA (Cherenkov Telescope Array) collaboration, we will eventually participate in the definition of a universal data format for gamma-ray astronomy (in the view of CTA).

5 Conclusions

The MAGIC Data Center based at PIC has been providing quality service to the MAGIC collaboration for several years now, exploiting when possible the extra resources that a GRID-based infrastructure implies.

The success of our work so far makes us push further our possibilities. We keep improving the access to the data by the MAGIC analyzers, and we work on making it more transparent for the interested astrophysicists outside the collaboration.

References

[1] Albert, J., et al. 2008, Nucl. Instr. Meth. A, 594, 407

[2] Moralejo, A., et al. 2009, Proceedings of the 31st ICRC, Lodz

[3] Oya, I., et al. 2009, Highlights of Spanish Astrophysics V, CD-ROM

[4] Reichardt, I., et al. 2009, Proceedings of the 31st ICRC, Lodz