

Data Management and Processing for the MAGIC Telescope

D. Dorner^a, K. Berger^a, T. Bretz^a and M. Gaug^b on behalf of the MAGIC Collaboration^c

(a) *Institut für Theoretische Physik und Astrophysik, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany*

(b) *Institut de Física d'Altes Energies (IFAE), 08193 Bellaterra, Barcelona, Spain*

(c) *Updated collaborators list at: <http://magic.mppmu.mpg.de/collaboration/members/index.html>*

Presenter: Markus Meyer (dorner@astro.uni-wuerzburg.de), ger-dorner-D-abs1-og27-poster

Every observing night the MAGIC Telescope is producing up to 170 GB of scientific data and this number will increase with the installation of 2 GHz FADCs. To provide a stable analysis of all the data an automatic data processing is essential.

For MAGIC a concept has been realised, in which an automatic, failsafe preprocessing of all data is ensured. The status of the analysis can be queried at any time from a database. The flexibility of the concept makes it easy to add new steps at any point of the analysis chain.

1. Introduction

With its 577 pixel camera and its 30 FADC slices per pixel the MAGIC Telescope produces up to 170 GB of data each night stored in about 500 files. To transfer this large amount of data from the Canary island La Palma, where MAGIC is situated at a height of 2200 m a.s.l., there are two procedures: Small amounts of data like the run information and the output of the online analysis are transferred daily via internet. The large amounts of raw data are then transferred on tape via mail every one or two weeks. Future plans to exchange the current FADCs by 2 GHz FADCs and to store 80 FADC slices promise even larger amounts of data. In addition the data volume will grow with the beginning of operation of the second telescope [1].

To handle such large amounts of data an automatic procedure is mandatory. A concept and its realisation for the MAGIC project will be presented in this paper.

2. A Concept of Data Management and Processing

To deal with more than 500 files and more than 150 GB per night the data processing procedure has to fulfil two important needs: It has to be failsafe to guarantee that all data is processed and automatic to make sure that all data is treated the same way and the analysis results are consistent. For MAGIC an automation concept has been developed, which meets not only these requirements but offers also many other useful features. A systematic scheme of the dataflow and the key elements of the concept are shown in Fig. 1. In the following the concept, which is based on Shell, MySQL and C++, will be explained in detail.

One of the core elements of the structure is a MySQL database, which consists of different kinds of tables. On the one hand there are tables which contain any kind of information: run information (i.e. information about each run like for example source name, starttime, etc.), information from the runbook (i.e. comments from the shiftcrew in the electronic runbook), calibration results (i.e. conversion factors [pHe/FADC counts], bad pixels, etc.) and more. On the other hand there are the status tables: for each step in the dataflow and the analysis there is a field in one of the status tables. To meet the given conditions there are several status tables: some steps have to be done per run, therefore a table on run base is needed. Other steps are done per night, others per dataset (to analyse more data of one source together runs are grouped to datasets, as shown in Fig. 3). Adapted to the needs of the experiment there can be several status tables. If a certain step has been executed successfully, the time is entered in the corresponding field. Therefore it is not only known, whether the step has already been done, but also when it has been done. The users thus have the information with which software version the

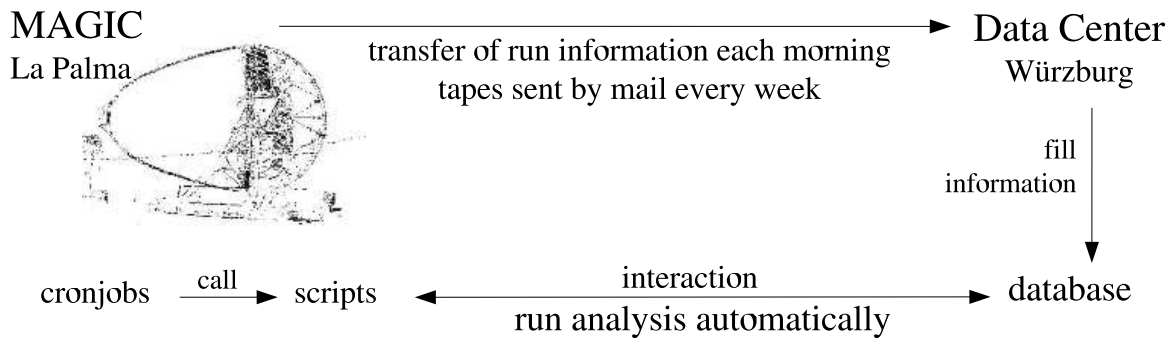


Figure 1. Schematic diagram of the automation concept for the MAGIC Telescope. Small files like run information are transferred every day via internet from La Palma to the datacenter. The big amounts of raw data are sent on tape by mail. From the arriving files all useful information is extracted and filled into the database. The automatic analysis is done with scripts that are called by cronjobs. Details on the interaction between the scripts and the database can be seen in Fig. 2.

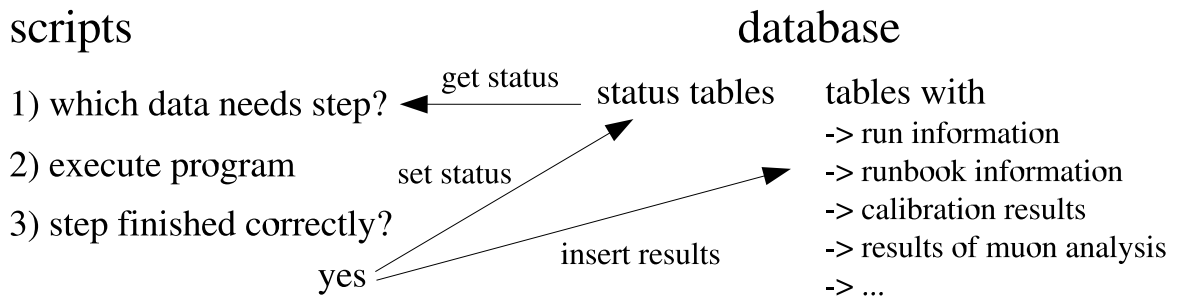


Figure 2. Scheme of the interaction between the scripts and the database. For a step, that has to be executed, the script performs the following steps: 1) query from the database, for which data the step is needed, 2) execute the program for this data and 3) if the step has finished without error, insert the new status into the database and insert the results of this analysis step into the appropriate table.

data has been analysed. If a step has to be repeated, for example, when there is an updated software version, this can easily be done, by removing the time from the corresponding field. As the time of execution is known, it is also possible to rerun a step only, if the data was analysed with a certain software version, by sending an appropriate query to the sql server.

For MAGIC the following implementation has been made (Figs. 1, 3): In a daily transfer run information is copied from the telescope site in La Palma to the datacenter in Würzburg. All usefull information is immediately filled into the database. These steps are done on a nightly base, whereas for example the datacheck is done run wise. The runs are grouped to sequences before calibration [7] and the calculation of image parameters [2] is done. Thus the sequence status table contains for example fields for calibration and image parameter calculation. How the automation works in this case is shown in Fig. 3. To analyse big amounts of data the sequences are grouped to datasets and the remaining steps are done on dataset base. In a structure as explained above the status of each run, sequence or dataset is stored in the database and can be queried either from an automatic script or from a user via web interface so that the status of the analysis can be followed easily.

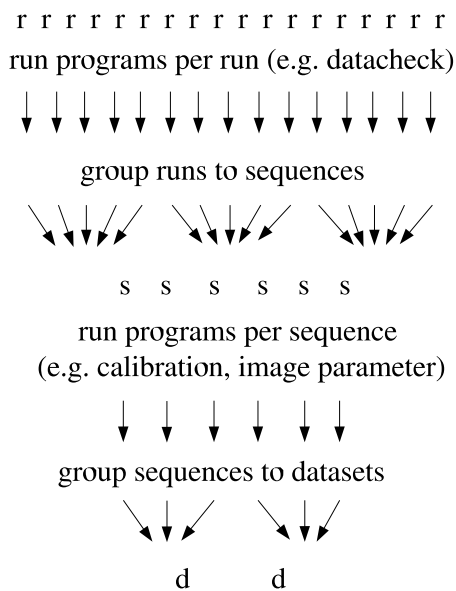


Figure 3. Schematic overview of the data structure in MAGIC: After some programs (e.g. datacheck) have been executed run wise, the runs are grouped to sequences. Some steps of the analysis are done sequence wise, like for example the calibration [7] and the calculation of the image parameters [2]. To analyse more data of one source together the sequences are grouped to datasets on which the rest of the analysis is performed.

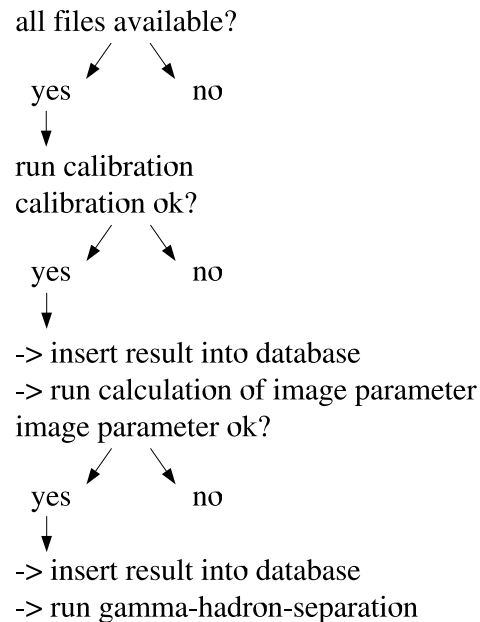


Figure 4. Example for the automatic procedure for the analysis of a Cherenkov telescope: If all files of a sequence of runs are available in the datacenter, the calibration is done for this data. If the calibration does not return any error, the result is filled into the database and the calculation of the image parameters is carried out. If this was successful the results are inserted into the appropriate tables in the database and the gamma-hadron separation [6] is performed. How the interaction between the scripts and the database looks like for the single steps, is shown in Fig. 2.

Based on this database structure the automation can be realised with simple shell scripts, one for each step in the analysis chain. Like shown in Fig. 2 the script queries from the database for which data a certain step has to be done and executes the program for this data. After the step has been completed without returning any error, the new status for the data is inserted into the database by inserting the current time into the corresponding field. Thus it is guaranteed that the data processing is fail-safe even in cases of computer problems, power failure etc. If a step could not be executed completely, it is retried the next time, when the script is started automatically. Only if data cannot be treated with the standard software, manual interaction is needed.

It is important, that all steps are performed in the correct order. For example the calibration can be done only, if all files are available. Therefore the check for disposability of all needed files has to be performed first. But the calibration also influences the next steps: the calculation of the image parameters and the filling of the calibration results into the database need the successful completion of the previous step, the calibration (see Fig. 4). These interdependencies of the steps are stored in a setup file, from where the scripts retrieve them.

By calling the scripts with cronjobs (see Fig. 1) no manual interaction is needed. As soon as the data is on disk, it is processed automatically.

By calling the scripts via cronjob it is very easy to run all analysis programs, that do not need any manual interaction, automatically. Accordingly all data is processed automatically, as soon as the files have been copied from tape. The concept also suits well for the usage with a queuing system like for example condor.

In the implementation for MAGIC the scripts work in interaction with the **M**agic **A**nalysis and **R**econstruction Software (MARS, [3, 4, 5]). Like this it is easy to implement the steps of the standard analysis chain [2] in the automation concept.

Due to the automation it is possible to obtain first results [2] quickly for all data. With the various tools of MARS the database also gives the possibility to do datachecks and long-term studies of quality parameters. For example the results of the muon analysis are stored in the database and thus the evolution of parameters like the point spread function could be examined [8].

3. Flexibility

A very important point is the flexibility of this automation concept. In an evolving and expanding experiment like MAGIC it is important that the concept is adaptable to new software.

The automation for MAGIC has been designed such, that it is possible to add new steps in the analysis to the whole structure easily. Only three small actions are necessary to implement a new step: 1) add a new field in the corresponding status table in the database, 2) adapt the setupfile with the dependencies of the step, and 3) write new script, that calls the program, and add it in the cronjob.

4. Conclusions

Based on the data stream produced by the MAGIC Telescope, we have shown, that it is possible to handle large amounts of data and files automatically with the presented concept. The interaction with a database provides the possibility to check the quality and status of the data at any time. By executing the programs of MARS the implemented structure delivers a quick look into the data with a robust analysis and first results, as soon as the data arrives in the datacenter. Due to its flexibility the concept can be easily expanded and adapted for the next development stage of the MAGIC experiment, but also for other future experiments.

5. Acknowledgements

We acknowledge support by the german Federal Ministry of Education and Research (BMBF, 05 CMOMG1/3) and the Astrophysics Institute of the Canary Islands (IAC).

References

- [1] M. Teshima et al., these proceedings
- [2] T. Bretz et al., these proceedings
- [3] T. Bretz et al., 28th ICRC, Tsukuba (2003)
- [4] T. Bretz et al., AIP Proceedings 745, p.730
- [5] T. Bretz, D. Dorner, Toward a Network of Atmospheric Cherenkov Detectors VII, Palaiseau (2005)
- [6] B.Riegel, T. Bretz et al., these proceedings
- [7] M. Gaug et al., these proceedings
- [8] F. Goebel, K. Mase, M. Meyer et al., these proceedings