*Proceedings of the 30th International Cosmic Ray Conference*
Rogelio Caballero, Juan Carlos D'Olivo, Gustavo Medina-Tanco,
Lukas Nellen, Federico A. Sánchez, José F. Valdés-Galicia (eds.)
Universidad Nacional Autónoma de México,
Mexico City, Mexico, 2008

Vol. 5 (HE part 2), pages 1057–1060



30TH INTERNATIONAL COSMIC RAY CONFERENCE

# A GRID approach to ARGO-YBJ experiment data transfer and processing.

C.STANESCU[1], A.BUDANO[1], P.CELIO[1], S.CELLINI[1], F.GALEAZZI[1], Y.Q.GUO[2], S.MARI[1], L.WANG[2], X.M.ZHANG[2] FOR ARGO-YBJ COLLABORATION
[1]*Dipartimento di Fisica, Universita' Roma Tre and INFN Roma Tre, Roma, Italy*
[2]*IHEP, Beijing, China*
*stanescu@roma3.infn.it*

**Abstract**: Some aspects of the cosmic ray astronomy require the access and the processing of the data in the shortest possible time. We implemented a data files moving system, based on GRID tools and services, to automatically transfer the files from the high altitude ARGO-Yangbajing Laboratory in Tibet to the Storage Elements at the processing sites in IHEP-Beijing (China) and CNAF-Bologna (Italy). We also describe the GRID approach to an unified job submission and processing system and the mirroring of the ARGO data files catalogs. This new approach allows our communities to cooperate more efficiently in data analysis, to share the available resources and to realize at the same time the backup of the data.

## Introduction

The Chinese-Italian cosmic-ray telescope ARGO-YBJ in Tibet has been completely installed and put in data acquisition with the full carpet installed. It is based on a single layer of RPC chambers organized in groups of 12 (cluster) for data acquisition. This layer is composed by 1848 chambers (with 18480 pads) plus an external ring made by 240 chambers. The detector is optimized for small air showers detection for studies in gamma astronomy, gamma-ray bursts, the Anti-p/p ratio, the primary proton spectrum, etc. The observation of low energy phenomena is possible due to the location of the experimental apparatus at high altitude (4300 m above see level) and to the high active surface (> 92 %). The trigger is based on pad multiplicity, that corresponds to a data acquisition rate of around 4 KHz of events, producing after data compression around 2.5 Mbyte/s of experimental raw data.

### Resource evaluation

The data taking is organized in RUNs composed by files of around 1 Gbyte each. Before writing the data to the files, the raw data events are submitted to a data cleaning procedure to eliminate the redundant information. The average dimension of an event is around 650 bytes and will grow in the future, when data from the RPC outer ring will be added to the DAQ system. A raw data file contains more than 1.7 millions of events. Taking into account a duty cycle of 85%, in an year we are going to register more than $1.3 * 10^{11}$ events. The reconstruction procedures requires around 200 KSPECint2000 of computing power, considering a (low) security factor of 50% for incertitude and reprocessing. At the level of the Monte-Carlo production, we foresee the use of just a small fraction of simulated showers in respect to the experimental data taking and the reuse of the simulated showers. We estimate a request of other 200 KSPECint2000 for the calculation related to MonteCarlo production.

## A GRID approach

The experimental site at Yangbajing provides no such facility as a computing center. The computing resources available at the site allow only for some limited data processing and data storage.
Hence the data collected by the experiment need to be moved and analyzed elsewhere: the collaboration relies on two computing centers, one in IHEP-Beijng in China, and the other in CNAF-

Bologna in Italy. Usually the data collected by the experiment are written to tape and sent to the main computing centers, where the tapes are read and the data copied to disk. The consequence is that the data are available to the collaboration after a delay of the order of weeks or greater. This not only affects the physics analysis but also has an impact on the control of the experiment, since some subtle effects may be revealed only after a full data analysis. A 155 Mbs network link was set between Yangbajing experimental site in Tibet and IHEP in Beijing, allowing us to carry out the raw data via network.

However we need a more organic approach to data transfer and processing and to facilitate a closer collaboration between Italian and Chinese data analysis groups. The GRID technology offers a wide variety of possibilities to build an integrated system for data moving, data processing, sharing of the resources and information, enforcement of common policies, redundancy, etc. We formed a group of study and integrate our work and efforts in a wider frame of an Euroean Program aiming to a closer collaboration between and EU and China. In this project aiming to extend the GRID infrastructure for e-Science to China and named EUChinaGRID, ARGO-YBJ experiment is one of the application package inside the project (see [1]). Some working package inside the project was dedicated to the connectivity between EU and China and stimulated the activation and the use of two different network connections to China : TEIN2 [2] and ORIENT [3], both at 2.5 GBPS bandwidth.

## Data Moving

We started to develop the so called “Data Mover Application” to transfer data from the ARGO-YBJ laboratory to the collaboration computing centers using gLITE Grid services. The “Data Mover” application is based on four Grid services: the Storage Element, the File Transfer Service, the Logical File Catalog and the User Interface.

The File Transfer Service (FTS) is the component that permits to move in a controlled way the data from an Storage Element SE to another.

The FTS service works with "channels" that connect the SEs. A channel is a named uni-directional logical connection from one SE to another and it is configurable in terms of bandwidth, number of streams, access policies, etc. Transfer of one file or of a group of files is called a “job”. FTS jobs are processed asynchronously: upon submission a job identifier is returned, which can be used at any time to query the status of the transfer.

The Logical File Catalog (LFC) permits to the GRID users to assign a logical name to a physical file present on a SE. The association is one-to-many, a logical name can point to several physical copies of the same file (“replicas”).

The User Interface is the gateway to the Grid, where users are authenticated and authorized to use the gLITE Grid services.

The design and the implementation of a data movin system for ARGO-YBJ was based on the schema presented in Fig 1., where the arrows indicate the FTS channels

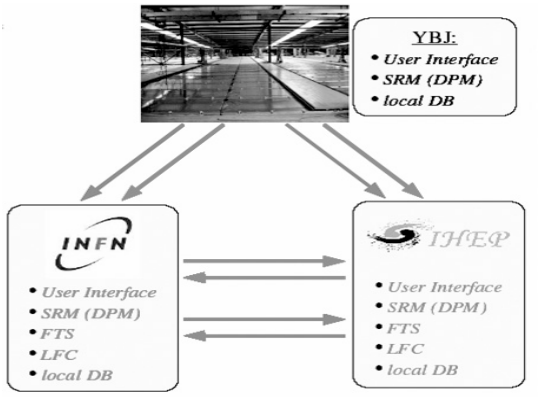


Figure 1: Data Mover Organization

The system was built with a certain degree of redundancy, using more FTS servers and defining many channels for the same destination.

At the DAQ level the experimental data are sent to the local storage system and routinely migrated to the SE by a program running in crontab. Since the SE and the farm machine share the same disk the migration involves no data copying, rather only the metadata information stored in the SE database is updated. As soon as a run has been successfully migrated, a flag is set in the local DAQ database.

At this point the “Data Mover”' (DM) application starts. The architecture of this application can be split in three sub-programs: transfer of the data

from YBJ to one of the computing centers, synchronization of the catalogues between the two computing centers and garbage collection at YBJ.
The data transfer procedure selects runs for which the migrated flag is set, prepares the list of relevant files, picks the FTS server and channel to be used based on their availability, and submits the files for transfer.
As far as the FTS server and channel choice are concerned, the first available FTS server is contacted, and for such server the first working channel from YBJ to one of the computing centers is used ; in case of problems the other FTS server and/or channel are tried.
After a run has been queued for transfer, the FTS server used and the id of the transfer are stored in the Data Mover dedicated database for monitoring purpose. The application periodically checks the state for every transfer and when all data files of a run have been copied they are registered in the LFC catalog ; the Data Mover database is updated accordingly.

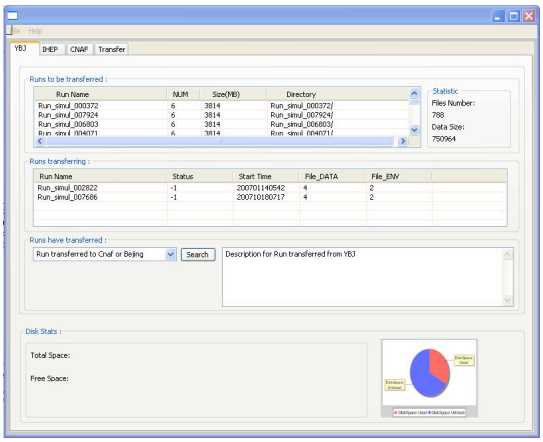

Figure 2: GUI for Data Mover

The synchronization process runs asynchronously at each site. It queries the LFC catalog of the other site and tries to find entries which are not in the local catalog. Missing entries are selected for transfer using the first available FTS server/ channel pair, as above. When the transfer is complete the LFC local catalog is updated: furthermore, the copy of the file at the other site is registered locally as a replica and the local copy is registered as a replica at the other site.

The garbage collector is responsible for cleaning up the buffer disk at YBJ. All files that have been recently transferred are checked and the files for which two distinct physical copies are found in the LFC catalogues are removed from the buffer disk. If there is still need for disk space, the garbage collector starts to copy the files to tape. Upon successful copy to tape, the files are deleted. The tapes will then be sent to Italy, read and the files stored in the SE and registered in the LFC catalogue.
For the "Data Mover" application test we used a SE in INFN Roma Tre to simulate the YBJ site. The whole testbed was based on Roma Tre, CNAF, IHEP sites and used several files of about 100 MB, stored in the Roma Tre SE. The test was successful. The application showed a very high reliability, and always granted that at least two copies of the data files existed at any time [4].
For better transfer performance between Italy and China sites some aspects of routing policy have to be improved. A graphical interface to monitor continuously the status of the data transfer was also developed (see Fig.2).

## Sharing the resources

The GRID technology gives us the possibility to share the computing resources and to obtain the same results in less time. The reconstruction process should work in more controlled environments meanwhile the MonteCarlo production and data analysis could use wider and more fragmented GRID resources. So we defined a simmetrical architecture with two main GRID production sites for raw data reconstruction, both storing a copy of the raw and reconstructed data, continuously aligned. The porting of ARGO-YBJ production in GRID environment was accomplished in several steps:
First we defined an ARGO Virtual Organization, the roles and an ARGO VOMS package to be installed in all the sites supporting ARGO. The mutual acceptance of CA certificates was also established.
After we tested the compliance of ARGO-YBJ reconstruction and Monte Carlo software with GRID environment and ask to software managers to install the ARGO official software on CE repository.

The scripts used for ARGO reconstruction and Monte Carlo production was modified in order to work in GRID, with the job submission procedure translated in JDL scripting language.
A special care was devoted to the modification of the ARGO production database. The data files can be retrieved now trough their LFC logical name and we can be informed about their replicas. The ARGO production database is a key point for ARGO GRID production, being the only point with instantly updated information about raw data files in reconstructing phase.
The GRID architecture we are going to use for ARGO production is indicated in Fig.3. Each production site will keep a copy of the raw data and of the reconstructed data, an LFC data catalog, a BDII information systems and the user interfaces. There should be only one active

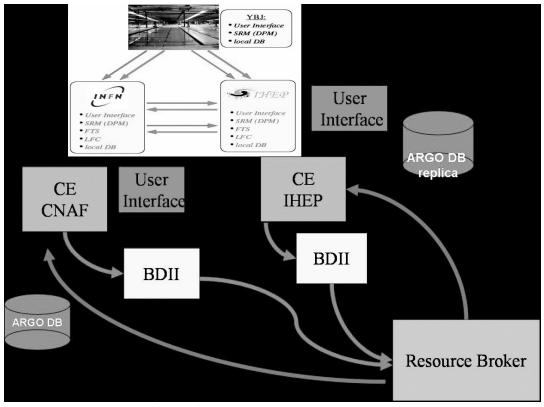

Figure 3: ARGO GRID job submission system.

production database that will be inquired to discover the raw data files to be processed and the request for computing resources will be forwarded to a Resource Broker RB. The jobs will be submitted according to the availability of computing resources and automatically the data files will be read in input and written in output on the local SE. We are using at CNAF a SE of the type SRM/CASTOR and at IHEP SRM/dCACHE.
One the job is finished the database will be upgraded and the reconstructed events files will be copied to the other computing center SE through the same procedure based on crontab, used for Data Mover (see above). For safety reasons the production database will be mirrored allowing the users at different sites to enquire and select data for physics analysis independently of the status of the network links.

## Conclusions

The procedures for Data Mover was tested on a testbed including the INFN Roma Tre, CNAF and IHEP sites. Some minor configuration problems for FTS servers was detected and solved. Some performance problems still persist during the files transfer and should be solved by the network managers through better routing procedure. The last step foresees the installation at Yangbajing site, in Tibet, of a SE as the destination point for raw data collection and will be done simultaneously to the foreseen DAQ hardware upgrade this summer. The automatic procedure to transfer the experimental data using the GRID middleware tools allows either a good control and monitoring of the operations and the fast availability of the data to the processing system, through the use of LFC catalogues.
The AGBO job submission was tested using different user interfaces and computing resources either in Italy or in China.

## References

[1] http://www.euchinagrid.org/
[2] http://www.tein2.net/
[3] http:/www.gridtoday.com/grid/642165.html/
[4] P.Celio et. al. The use of GRID tools for automatic file transfer of ARGO-YBJ experiment data, demo pres., EGEE User Forum, Manchester UK, May 2007