

# Application of the DIRAC framework to CTA: first evaluation

L Arrabito<sup>1</sup>, C Barbier<sup>2</sup>, R Graciani Diaz<sup>3</sup>, B Khélif<sup>4</sup>, N Komin<sup>2</sup>, G Lamanna<sup>2</sup>, C Lavalley<sup>1</sup>, T Le Flour<sup>2</sup>, JP Lenain<sup>5</sup>, A Lorca<sup>6</sup>, M Renaud<sup>1</sup>, M Sterzel<sup>7</sup>, T Szepieniec<sup>7</sup>, G Vasileiadis<sup>1</sup>, C Vuerli<sup>8</sup> for the CTA consortium

<sup>1</sup> Laboratoire Univers et Particules, Université de Montpellier II Place Eugène Bataillon - CC 72, CNRS/IN2P3, F-34095 Montpellier, France

<sup>2</sup> Laboratoire d'Annecy-le-Vieux de Physique des Particules, Université de Savoie, CNRS/IN2P3, F-74941 Annecy-le-Vieux, France

<sup>3</sup> University of Barcelona, Diagonal 647, ES-08028 Barcelona, Spain

<sup>4</sup> Laboratoire Leprince-Ringuet, Ecole Polytechnique, CNRS/IN2P3, F-91128 Palaiseau, France

<sup>5</sup> Landessternwarte, Knöigstuhl 12, D-69117 Heidelberg, Germany

<sup>6</sup> Universidad Complutense de Madrid, Avda. Complutense s/n, E-28040 Madrid, Spain

<sup>7</sup> Academic Computer Centre CYFRONET, ul. Nawojki 11, P.O. Box 38630-950 Krakow 61, Poland

<sup>8</sup> Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34143 Trieste, Italy

E-mail: [arrabito@in2p3.fr](mailto:arrabito@in2p3.fr)

**Abstract.** The Cherenkov Telescope Array (CTA) - an array of several tens of Cherenkov telescopes - is the next generation of ground-based instrument in the field of very high energy gamma-ray astronomy. The CTA observatory is expected to produce a main data stream for permanent storage of the order of 1-to-5 GB/s for about 1000 hours of observation per year, thus producing a total data volume of the order of several PB per year. The CPU time needed to calibrate and process one hour of data taking will be of the order of some thousands CPU hours with current technology. The high data rate of CTA, together with the large computing power requirements for Monte Carlo (MC) simulations, need dedicated computing resources. Massive MC simulations are needed to study the physics of cosmic-ray atmospheric showers as well as telescope response and performance for different detectors and layout configurations. Given these large storage and computing requirements, the Grid approach is well suited, and a vast number of MC simulations are already running on the European Grid Infrastructure (EGI). In order to optimize resource usage and to handle all production and future analysis activities in a coherent way, a high-level framework with advanced functionalities is desirable. For this purpose we have preliminarily evaluated the DIRAC framework for distributed computing and tested it for the CTA workload and data management systems. In this paper we present a possible implementation of a Distributed Computing Infrastructure (DCI) Computing Model for CTA as well as the benchmark test results of DIRAC.

## 1. Introduction

The current generation of Imaging Atmospheric Cherenkov Telescopes (IACTs) such as H.E.S.S., MAGIC and VERITAS, has opened in the last few years the realm of ground-based gamma-

ray astronomy in the so-called very-high-energy (VHE) domain, at energies above a few tens of GeV. The Cherenkov Telescope Array (CTA) project [1] is the next generation of IACT, operating in the HE/VHE (20 GeV-200 TeV) domain, and will consist of two arrays of 50-100 Cherenkov telescopes of 3 different sizes, located in each of the Northern and Southern hemispheres. The CTA consortium gathers about 800 scientists and engineers belonging to more than a hundred institutes world-wide and is currently under its Preparatory Phase (2011-2013). The construction of the full array is planned to be completed around 2018.

CTA will improve point-source sensitivity by an order of magnitude with respect to current IACTs. In addition, compared to any previous observations in this energy range, it will feature improved angular resolution (at the arcmin scale) and collection area. Moreover, thanks to a larger field of view and detection area, CTA will have an enhanced sky-survey capability. CTA will benefit from these high performances to explore different scientific cases, such as the investigation of the origin of the Galactic Cosmic Rays, the most violent astrophysical phenomena (e.g. AGNs, GRBs), fundamental physics and cosmology (indirect dark matter search, Lorentz Invariance violation investigation). Finally, CTA will be the first high-energy ground-based astronomical open observatory. This means that it will receive observation proposals from the whole astronomical community and a dedicated Science Data Center will have to provide transparent access to data, analysis tools and user training.

In this paper we report on a possible implementation of a Distributed Computing Infrastructure (DCI) to fulfill present and future requirements for Monte Carlo massive production and data analysis for the CTA observatory. In section 2, a preliminary estimation of the computing and storage needs is given, based on the extrapolation from current IACTs. In section 3 we present a preliminary and general view of the DCI implementation of the operational workflow of the CTA pipelines, which is one of the potential solutions currently under investigation within the CTA consortium. In section 4, the EGI solution and the tools (in particular the DIRAC system) adopted for the Monte Carlo simulations and investigated to implement more generally a possible DCI CTA Computing Model are presented. The current evaluation of DIRAC and its application to CTA real use cases in the context of MC simulations are provided in section 5. Section 6 is devoted to conclusions and perspectives for future work.

## 2. CTA computing needs

The CTA global data pipeline is constituted by five main steps bringing the scientific data up to their delivery to the users: acquisition, calibration, reconstruction, archive and distribution. With the CTA project being in its Preparatory Phase, it is difficult to have a precise estimate of the future computing needs, since some technical aspects, which have a critical impact on the data stream, data processing, archive and data dissemination, are not yet defined. For the time being, a very preliminary estimate of the computing needs for storage and data processing for the calibration and reconstruction of raw data is based on some extrapolations from the figure of merits of current running Cherenkov systems and on some expected optimizations in the software design. The relevant aspects of the CTA instrumental specifications are considered with respect to the currently running IACTs: the different number of telescopes and sizes, the larger number of pixels per camera, the mean trigger telescope multiplicity and the expected trigger-rate. The total expected data rate for a specific array configuration under investigation can vary from 0.3 to 4.0 GB/s for the South Array and from 0.1 to 1.3 GB/s for the Northern one, depending on the average number of time samples per pixel. These data rates translate to a per-year volume of data to be archived between 2 and 25 PB, assuming a 15% of observation time (duty cycle) over the whole year. Accounting for these points, we estimate that one hour of CTA data will require the off-line processing times given in Table 1. The expected computing ranges from a total computing time of 2k HS06 years of off-line processing time per year of observation to more than 7k HS06 years. This translates into hundreds of CPU-years according

to the available technology deployed in the EGI.

**Table 1.** Expected raw data rate and off-line computing time needed to process one hour of CTA data. HS06h is an hour of CPU used by a single processor with benchmark value HEP-SPEC06=1 [2].

	Raw data rate (GB/s)	Calibration (HS06h)	Reconstruction (HS06h)
CTA South Array	0.3-4.0	3000-20000	6000-20000
CTA North Array	0.1-1.3	1500-6000	3000-6000

These numbers represent only preliminary and indicative values of the expected computing needs which are affected at the moment by large uncertainty. The total computing time will be then even larger than the values given in the Table 1, when MC simulations and high-level data analysis will be considered. Moreover, the computing power for periodical data re-processing and the resources needed for the real-time data pre-calibration and reconstruction, are not taken into account in the above estimate. It is therefore expected that a DCI approach (see subsection 3.2) will be able to provide the computing resources necessary for science exploitation of CTA.

### 3. CTA data pipeline

The implementation of the information and computing infrastructures (i.e. the Computing Model, CM) for the global CTA data pipeline should take into account the specific CTA data flow, the operational framework of the observatory, and the open data access requirements of the user community.

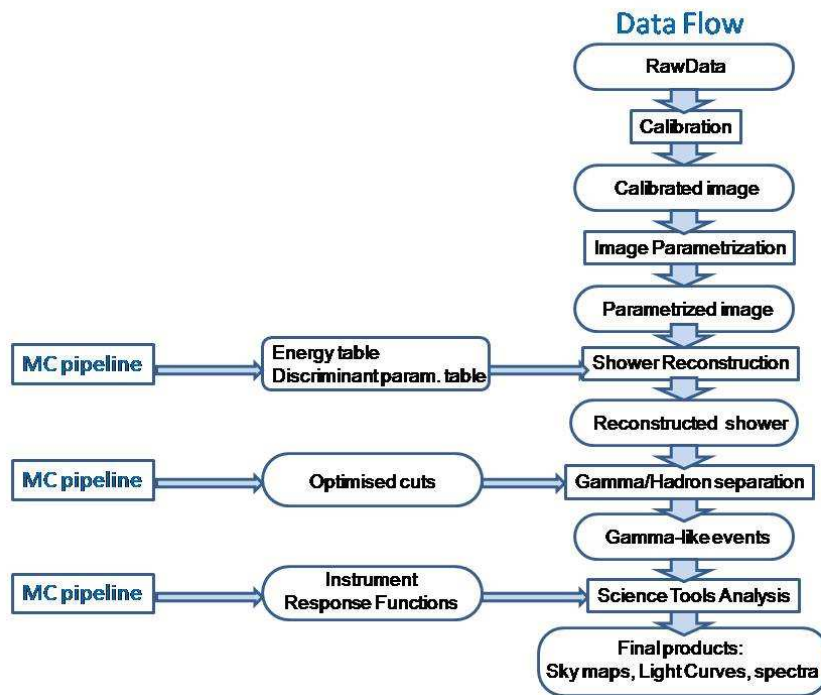
#### 3.1. CTA data flow

The CTA processing data flow (see Figure 1) can be decomposed in three stages:

- **Stage A:** corresponds to the raw data acquisition and it takes place at the Northern and Southern CTA sites, where the two arrays will be located. At this level, raw camera images, together with housekeeping and monitoring data, are collected from the cameras and auxiliary devices;
- **Stage B:** concerns the data processing, i.e. raw data calibration and all the steps of the data reconstruction (image parametrization, shower reconstruction, gamma/hadron separation);
- **Stage C:** deals with the so-called high level data (list of gamma-like candidates) and the related science analysis. At this level, science analysis tools are applied to the reconstructed data in order to produce astronomical data products, such as sky maps, spectra, light curves, etc.

The MC data pipeline is critical to both Stage B (calibration and reconstruction) and Stage C (instrument response function definition for high level products). Two more stages complete the overall data flow with the objective of data publishing:

- **Stage D:** concerns the archive of all data levels for both re-processing purposes as well as for coherent and efficient data distribution to CTA users and guest observers (scientist that submits an observing proposal);
- **Stage E:** data distribution through tools and solutions aimed at providing open access to all data, software and computing infrastructures.



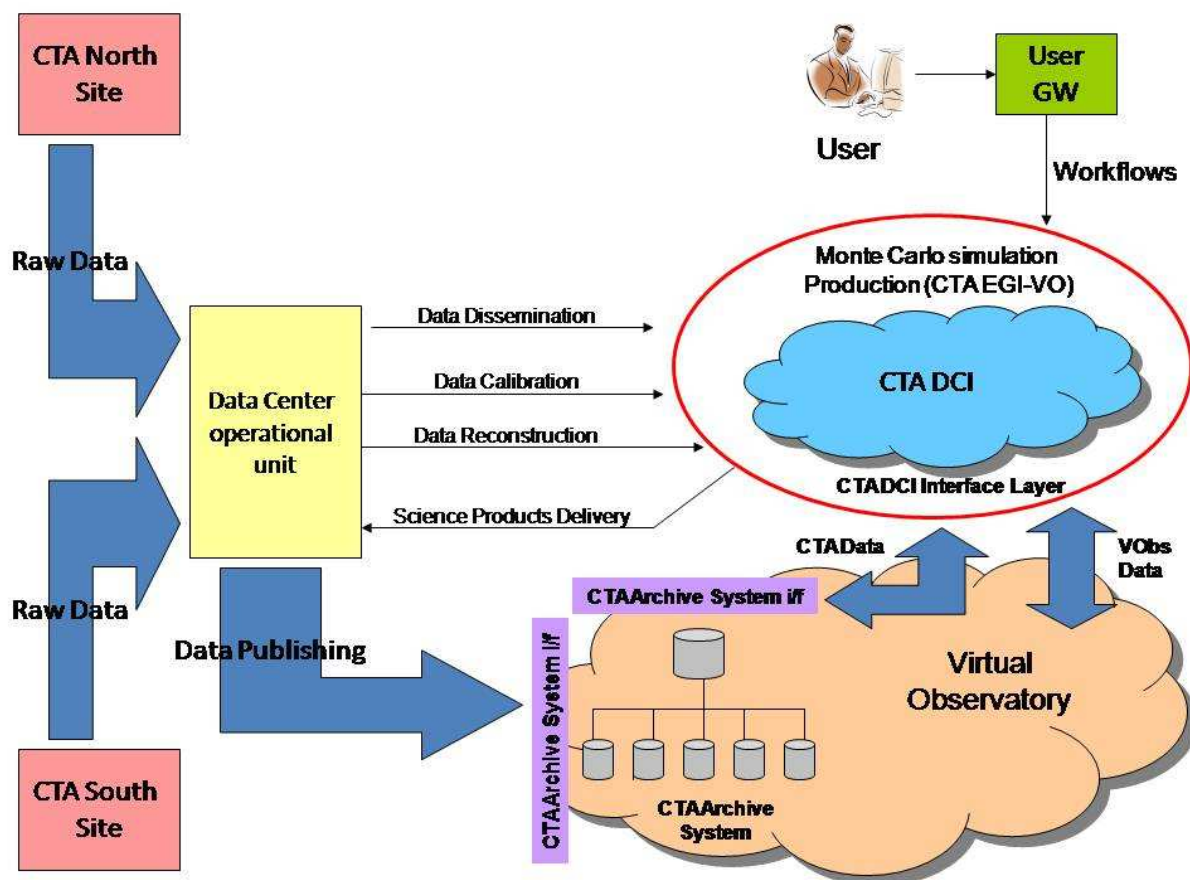
**Figure 1.** General overview of the CTA data flow from raw data to high level astronomical data. Round boxes represent data products, while rectangular ones represent data processing procedures. The role of the MC pipeline is also shown.

### 3.2. Proposal of a DCI approach for CTA

One of the main aims of the current CTA preparatory studies concerns the implementation of the operational workflow of the CTA pipelines and the corresponding CM. The CTA consortium has not yet finalized a final technical proposal, therefore the view presented in the following must be considered as preliminary. It corresponds to just one of the different potential solutions under examination within the CTA consortium. A global and not detailed overview of the main functional blocks of the proposed CM is shown in Figure 2. It interprets the main five stages of the data pipeline as described above and relates to a DCI approach, mainly inspired by the model applied to support the CTA Virtual Organization (vo.cta.in2p3.fr, CTA VO) [3] for the current MC massive production and data analysis. Its validity for the data pipeline has still to be examined in detail.

According to the data flow described in section 3.1, the operations related to Stage A will take place at the two CTA sites (North and South). Given the estimated data rates (see section 2), raw data will be stored in a local buffer and an on-site fast processing will proceed to check the data quality. Moreover, a first simplified reconstruction would also be performed on-site using the local computing facilities available at the two sites. Finally for Stage A, temporarily stored raw data will be delivered to a data center operational unit. More precisely, not only the raw data, but the whole CTA observatory data (raw, calibrated, high level and science data, plus apparatus housekeeping and monitoring data), will be stored at the data center, which will have the responsibility of the data custody.

Operations related to Stage B, i.e. data calibration and reconstruction, starting from raw data could take place with the support of a CTA DCI. Some motivations for the DCI approach are given at the end of this section. The final outputs of Stage B are the reconstructed data,



**Figure 2.** Overview of a possible functional CTA Computing Model based on a DCI approach.

which could be also stored within a DCI approach according to predefined replication and synchronization policies. Finally science analysis and related operations (Stage C in the data flow), based on end-user analysis workflows, could also take advantage of available CTA DCI if needed.

An Archive System will host and manage the access to all CTA data, including those which will be available to the whole astronomical community (Stage D). Public data must be made accessible not only through CTA specific tools, but also through the standard Virtual Observatory interfaces and protocols, and they will have to fulfill the Virtual Observatory standards. The last component of our CM is the so-called CTA Science Gateway (SG), which will most likely be the most important facility available to the end-users (Stage E). The SG will be conceived as the unique entry point to the CTA Archive for users and guest observers, as well as the access to an integrated science analysis system hiding the complexity of the e-infrastructure behind.

As previously seen, the CTA DCI plays a crucial role in our CM, since it handles the off-line data processing and science analysis. Considering the total computing power needed for baseline data processing (Stage B), as given in Table 1, and all further processing activities such as full processing (Stage B and Stage C), re-processing and multi-user analysis, a DCI approach is well suited. Moreover given the PB scale of the data produced every year by the full arrays, clear advantages would come from sharing e-infrastructures investment and responsibilities among

different international partners. The use of a Distributed Computing Infrastructure offers an optimal compromise between costs, development efforts, performance and risk minimization.

A DCI solution is also well suited to the requirements of the world-wide community of users who will access the CTA data. As already mentioned in section 1, CTA will be the first high-energy ground-based astronomical observatory. This means that after a proprietary period, data will be made available to the whole astronomical community. This community is very heterogeneous, as it is composed of both CTA consortium members and external scientists. They will access data at different stages of processing and will have different levels of Information and Communications Technology knowledge. All these types of users, although with different specific requirements, have some major common needs:

- a robust and flexible infrastructure;
- a secure and reliable connection;
- an efficient and homogenous access to computing and data resources;
- collaborate among geographically distributed working groups.

Given these requirements, a distributed approach is certainly appropriate since it offers high availability, reliability, safety and redundancy. Among the existing DCI solutions, scientific grids have already been exploited for several years by many scientific communities with similar high-level requirements.

#### **4. The CTA EGI Virtual Organization for Monte Carlo simulations**

All data processing at the current IACTs takes place in a single computing center, while at the scale of CTA, as seen in the previous section, a distributed approach would be more suitable. The high computing needs for Monte Carlo simulations in the Preparatory Phase of CTA, lead us to start exploiting the EGI precisely for large simulation productions (more than  $10^{11}$  proton, gamma, and electron induced showers are required for MC studies). A feasibility study of applications of Grid solutions for CTA is in progress within a dedicated CTA Computing Grid (CTACG) project [1]. The CTA Virtual Organization (CTA VO) was created in 2008 with the support of the IN2P3-LAPP computing center and in cooperation with the CC-IN2P3 computing center. Today the CTA VO is supported by 15 grid sites spread in 6 countries, with resources of the order of some thousands of available logical CPUs and more than 600 TB of storage. During the first years, we have developed tools, such as the EasiJob [4] infrastructure, which configures, submits and monitors MC productions. These tools have been using the gLite middleware [5] and GANGA [6] as backend for job submission. Typical MC productions consist of about 150000 jobs with peaks of more than a thousand simultaneously running jobs, producing about 200 TB of data.

More recently we realized the need to rely on a Workload Management System (WMS) to optimize the resource usage on the Grid and to keep a high job success rate. Large experience on Grid exploitation has been acquired in the past years by LHC communities, who all developed a WMS layer on top of the Grid middleware, precisely for the optimization reasons mentioned above. Among all these WMS solutions, we distinguished the DIRAC (Distributed Infrastructure with Remote Agent Control) system [7], which offers a very general framework to manage the distributed activities of a user community. In the next section, we will describe some of the most important functionalities of DIRAC, which make it a very promising solution for CTA. Web-based GUI tools to access the CTA VO resources, as well as to browse the data archives in the LFC catalogue have been developed within the CTA consortium (InSilicoLab [8]).

#### **5. Evaluation of DIRAC for the WMS on the Grid**

The DIRAC system, originally developed to support production activities of the LHCb experiment, is today a general solution that serves several communities (GISELA, ILC, Belle II,

CREATIS, SuperB, BES III, etc.). Even if DIRAC includes a Workload and a Data Management System (WMS and DMS), as well as many other components, like the DIRAC web portal, we will focus here on the WMS, which is the major component under evaluation for CTA.

The key feature of the DIRAC WMS is the implementation of the *Pilot Job* mechanism as a way to guarantee high success rate for user jobs (workloads) and community control on the policy for usage of the resources. In more detail, workloads are submitted to the DIRAC WMS and inserted in the DIRAC *Central Task Queue*. The presence of workloads in the DIRAC *Central Task Queue* triggers Pilot Jobs submission. Pilot Jobs are deployed on the Worker Nodes (WN) as regular grid jobs using the standard grid scheduling mechanism (gLite WMS or CREAM CE). Once started on the WN, the Pilot Job performs some checks of the environment and in case these tests succeed, the workload is *pulled* from the DIRAC *Central Task Queue*, executed and the results are returned to DIRAC. In case the environment checks fail or the Pilot Job gets aborted (for example because of a mis-configured site), only the Pilot Job is affected. The net result of this mechanism is a meaningful improvement on the workload success rate. Also, since the resources are retained by Pilot Jobs, the waiting time to get workload execution start is reduced.

Finally, another distinguishing feature of DIRAC is its capability to handle heterogeneous resources, such as Grids, Clouds and local clusters, providing a single entry point and hiding the underlying complexity. This feature is particularly important for a project like CTA, whose time scale is of the order of tens of years, as it should be able to adapt to the forthcoming evolutions of the distributed computing architectures, such as in the cloud case [10]. Being a high-level middleware solution, DIRAC is currently starting to provide an interface for third party application portals like gUSE/WS-pGRADE [11]. This is particularly interesting for the Science Gateway integration with DCI solution.

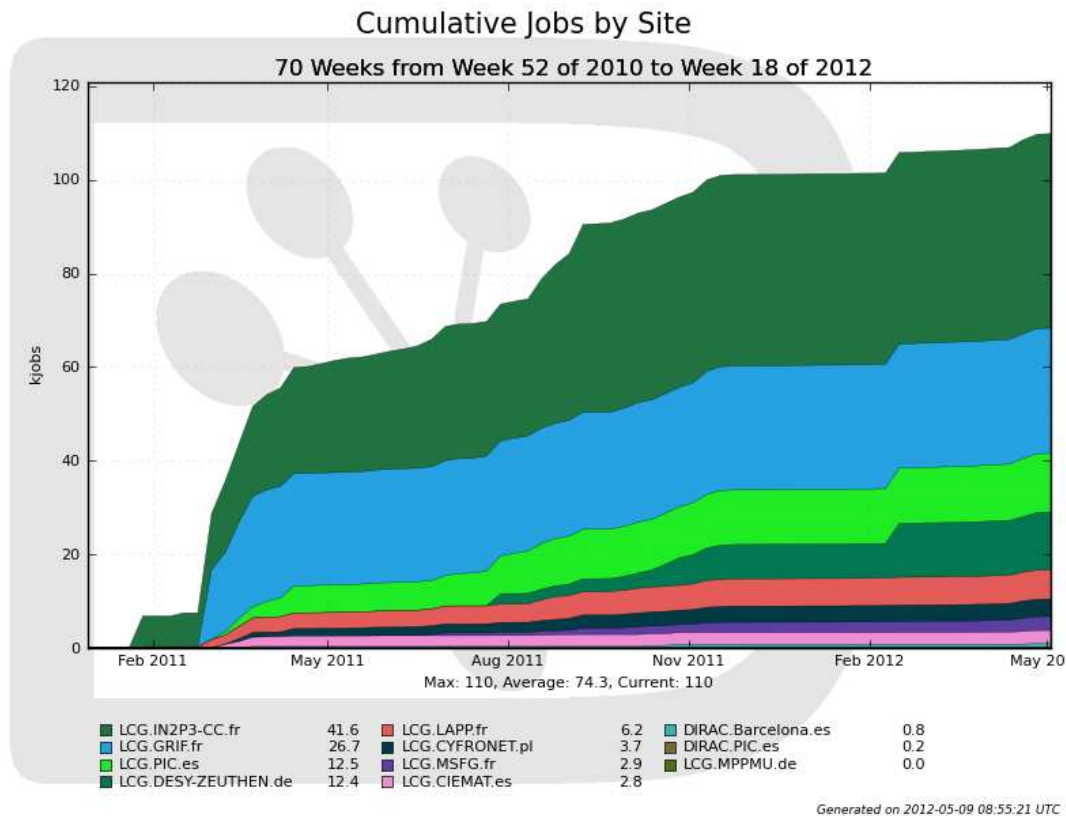
### 5.1. The CTA DIRAC instance

In order to make a realistic evaluation of the DIRAC system, it was felt necessary to have a DIRAC instance dedicated to the CTA VO. The installation of the DIRAC core services was implemented about one year ago at PIC in Barcelona and then one more at the CC-IN2P3 in Lyon, two Tier1s of the WLCG project, in order to improve the fault tolerance and redundancy of the DIRAC installation. These services include those associated to the WMS and DMS, as well as the DIRAC web portal and some other auxiliary components. Some of them are crucial for the global DIRAC operation. As an example, the Configuration System handles all the parameters necessary to DIRAC operations, and some redundancy is highly desirable. Recently DESY-Zeuthen has also offered to add another DIRAC relay point in order to make our DIRAC installation more robust.

### 5.2. CTA workflow definition within DIRAC

Once the DIRAC instance was configured and it started accepting and handling job requests, we started to develop some CTA specific workflows in DIRAC. In particular, we started to build workflows handling the full reconstruction pipeline, which can already be applied to CTA MC data. Since the reconstruction CTA software is not yet ready, we decided to implement workflows for the software in use by the current gamma-ray astrophysics experiments, and in particular the H.E.S.S. software. This software is also currently used for MC simulations, aiming to define the optimal CTA configurations for electronics, array layout, etc.

The whole reconstruction pipeline is quite complex, since it includes many interdependent steps, linked by their input/output. However, the main steps will remain valid for CTA, since they reflect the common concepts of any reconstruction pipeline of IACT data. The way how a DIRAC workflow is built is by connecting several modules that can be executed on a single job or split as convenient. In our case, each module corresponds to the call of a single application



**Figure 3.** Cumulative number of executed jobs through the DIRAC system during last year. The different colors are associated to the different Grid sites where jobs were executed. The major features of this step-like profile are associated with basic tests of the DIRAC framework (February 2011), generic CTA simulations productions (March-April 2011) and productions running CTA-like reconstruction pipeline (September-October 2011).

of the reconstruction pipeline, where some outputs of a previous module serve as input to the subsequent one. This work is achieved for the H.E.S.S. reconstruction pipeline, with first mini productions sent to the Grid and successfully executed. The future plan is to apply such DIRAC workflows to the next massive CTA MC production campaign.

In order to have the complete chain of reconstruction and end-user analysis within DIRAC, we also integrated two potential CTA science analysis tools, within DIRAC modules. A first exploitation of these modules has already been done in the context of the CTA scientific case studies to produce new results. At the same time, some work has started to connect the CTA DIRAC installation to the InSilicoLab application portal from CyFRONET, that already supports some of these CTA analysis applications.

In parallel to the workflow definition, we are also working on the handling of metadata linked to our reconstruction workflows. These metadata are essential to the coherence of the whole pipeline, in particular because they allow to clearly qualify the outputs of each step, so that different datasets can be selected as an input for further steps, according to some specific criteria (provenance and quality for example). With this purpose, we will use the DIRAC File Catalog facility, which is based on a relational Data Base schema and for which DIRAC provides python APIs that can easily be used from workflow scripts. In the context of the DIRAC evaluation, more than a hundred thousand jobs were handled by DIRAC and executed on the Grid during

last year (see Figure 3). Finally, concerning performances evaluation, the production experience of on-going experiments with much higher computing requirements than CTA [9], has proven that their DIRAC installation shows no scalability issues up to over 40 thousand concurrent jobs and sustained job rates over 100 thousand jobs per day, these limits being due to the availability of resources. Therefore, in order to process raw data in real time, taking into account the processing needs given in Table 1 and assuming an average job duration of 24 hours, we would need to handle a job rate of about 200-1000 jobs per day, well within the DIRAC system capability.

## 6. Conclusions and perspectives

In order to handle the large amount of data that will be collected by the CTA instrument, the DCI approach is considered as a key element of the CTA Computing Model. Motivated by the high computing needs of the MC simulation, we started the exploitation of the EGI in 2008 with the creation of the CTA Virtual Organization. In order to explore the available solutions to optimize the resource usage, an extensive evaluation of the DIRAC framework has been running since almost one year and the results obtained so far are very promising in terms of functionalities. On a medium-term scale, massive exploitation of the DIRAC system is foreseen during the next MC production campaigns, which will serve as test bench to validate our preliminary results. Further developments will aim to interface DIRAC with a prototype of the CTA Science Gateway, which is another essential piece of the CTA Computing Model.

## Acknowledgments

We gratefully acknowledge support from the following agencies and organisations: Ministerio de Ciencia, Tecnología e Innovación Productiva (MinCyT), Comisión Nacional de Energía Atómica (CNEA) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) Argentina; State Committee of Science of Armenia; Ministry for Research, CNRS-INSU and CNRS-IN2P3, Irfu-CEA, ANR, France; Max Planck Society, BMBF, DESY, Helmholtz Association, Germany; MIUR, Italy; Netherlands Research School for Astronomy (NOVA), Netherlands Organization for Scientific Research (NWO); Ministry of Science and Higher Education and the National Centre for Research and Development, Poland; MICINN support through the National R+D+I, CDTI funding plans and the CPAN and MultiDark Consolider-Ingenio 2010 programme, Spain; Swedish Research Council, Royal Swedish Academy of Sciences financed, Sweden; Swiss National Science Foundation (SNSF), Switzerland; Leverhulme Trust, Royal Society, Science and Technologies Facilities Council, Durham University, UK; National Science Foundation, Department of Energy, Argonne National Laboratory, University of California, University of Chicago, Iowa State University, Institute for Nuclear and Particle Astrophysics (INPAC-MRPI program), Washington University McDonnell Center for the Space Sciences, USA. The presented work has been partially financed by *Comisión Interministerial de Ciencia y Tecnología (CICYT)* (project FPA2010-22056-C06-02).

## References

- [1] Actis M et al. (CTA Consortium) 2011 *Experimental Astronomy* **32** 193
- [2] <http://hep.caspr.it/benchmarks/doku.php>
- [3] Barnacka A et al. 2012 *Lecture Notes in Computer Science* **7136/2012** 301-313
- [4] Komin N et al., <http://hal.archives-ouvertes.fr/hal-00653017>
- [5] <http://glite.cern.ch/>
- [6] Moscicki J T et al. 2009 *CoPhC* **180** 2303
- [7] Tsaregorodtsev A et al. 2008 *Journal of Physics : Conference Series* **119** 062048
- [8] InSilicoLab, <http://insilicolab.grid.cyfronet.pl/>
- [9] Graciani R *LHCb-PUB-2012-003* <http://cdsweb.cern.ch/record/1428064?ln=en>
- [10] Graciani R et al. 2012 *CHEP Parallel* 164 these proceedings

[11] Puig Navarro A et al. 2012 *CHEP Poster* 455 these proceedings