



CTA report

Luisa Arrabito, Ricardo Graciani, Claudia Lavalley



DIRAC User Workshop 29th – 31st October 2012, CPPM - Marseille

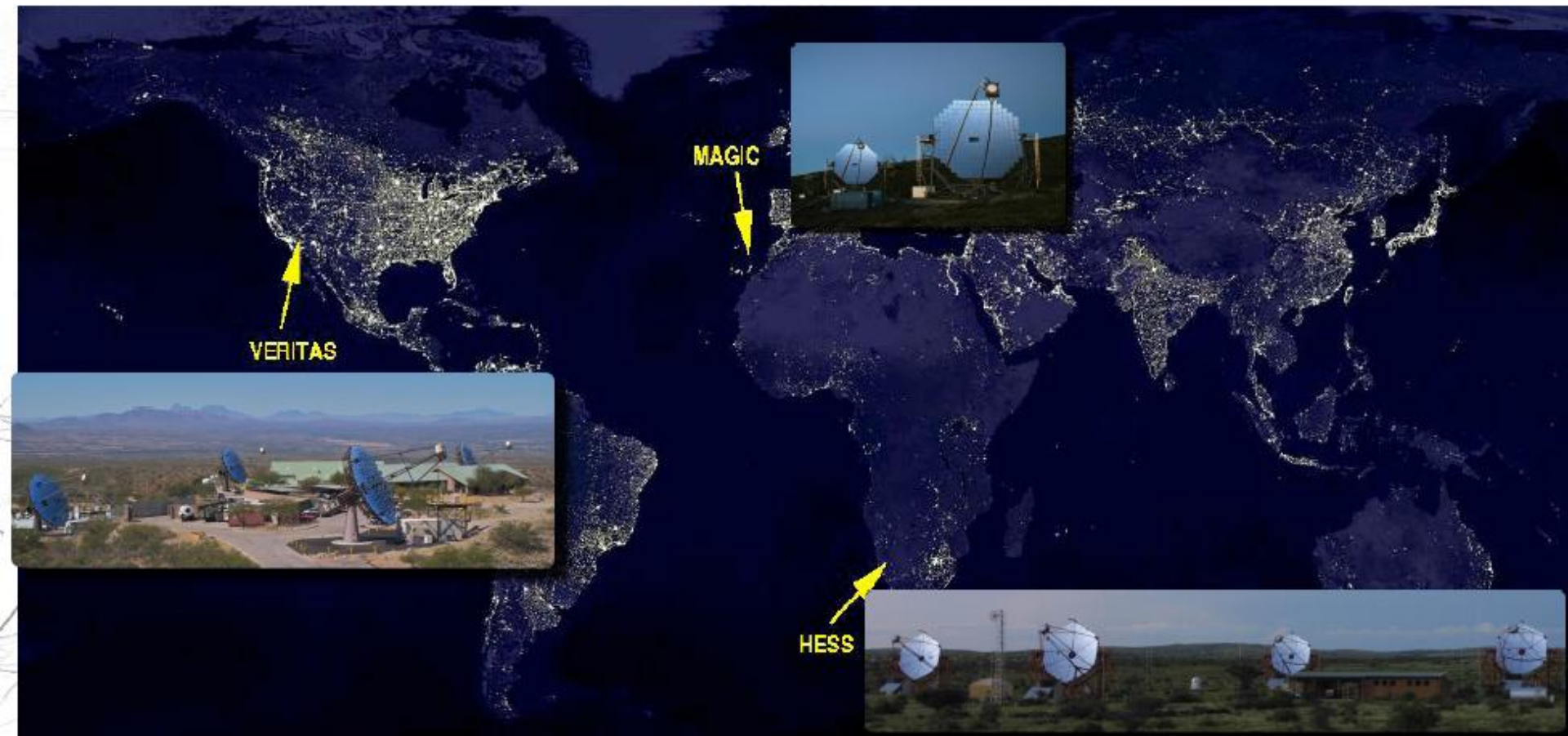
CTA report

Outlook:

- CTA project
- CTA computing
- MC simulations on the Grid
- Current use of DIRAC
- Conclusions and perspectives

Gamma-Ray Astronomy

- Astronomy at High Energy > 10 s GeV
- Currently **3 experiments with 2-5 Cherenkov Telescopes**: operational since 2003 (H.E.S.S.)
- Discovery of hundred's of sources in the last 8 years
- Observation of non-thermal universe: supernova remnants, pulsar wind nebulae, galaxies

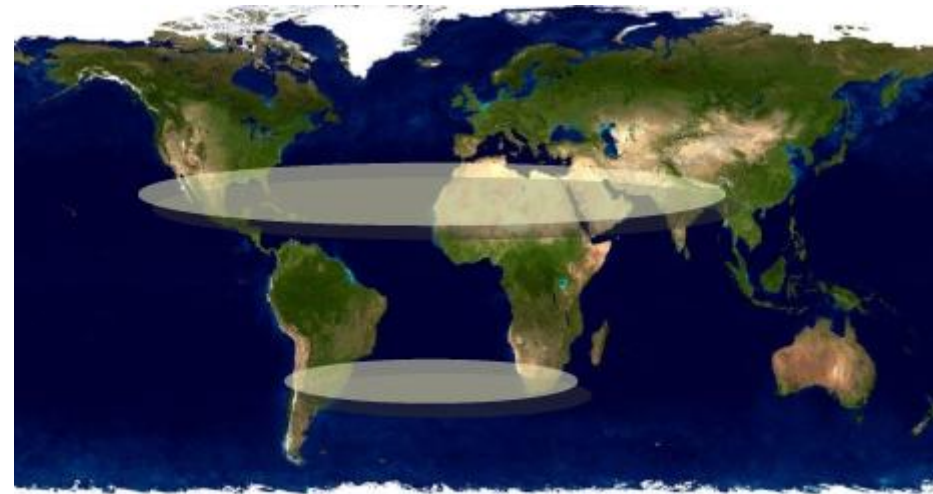


Next generation: Cherenkov Telescope Array (CTA)



Cherenkov Telescope Array

- 2 arrays of 50-100 Cherenkov Telescopes of 3 different sizes in each hemisphere
- Enhanced performances:
 - 10x sensitivity with respect to current experiments
 - Energy domain HE/VHE (20 GeV-200 TeV)
 - Improved angular resolution and collection area



Scientific goals

- Investigation of the **origin of the Galactic Cosmic Rays**
- Astrophysical phenomena (e.g. AGNs, GRBs)
- Fundamental physics and cosmology



Next generation: CTA

CTA Project

- Consortium of > 800 scientists in 25 countries
- Currently Preparatory Phase (2011-2013):
 - Find sites
 - Build prototypes of telescopes
 - Define optimal array layout
 - ***Prepare data pipeline and data distribution***
- The construction of the full array planned to be completed around 2018
- Operate as an observatory

CTA Computing tasks

Preparatory Phase:

MC simulation:

- Study of telescope and array design
- Algorithm optimization

Operation Phase:

- Raw data acquisition on remote site
- Data calibration and reconstruction
- Analysis and data publication
- MC simulation:
 - optimization of event selection cuts
 - generation of acceptance areas
 - energy reconstruction and resolution tables
 - point spread function

CTA Computing: Operation Phase

Computing needs

- **Data rate: 0.1 – 13 GB/s**
- **Total amount of Raw Data: 2 – 25 PB/ year**
- **Processing time** for 1 hour of observation (see Table below)

	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

Additional Computing power will be needed for:

- Periodical data re-processing
- Real-time data pre-calibration and reconstruction
- MC simulations
- High-level data analysis

-> Overall computing time still not estimated

- *No computing model yet:*

-> Distributed Computing approach will be explored for CTA operation

CTA Computing: MC simulations

MCsimulation:

- Production
- Analysis

Production

1) Air showers and Cherenkov light

- corsika sw [<http://www-ik.fzk.de/corsika/>]

2) Simulation of telescope response

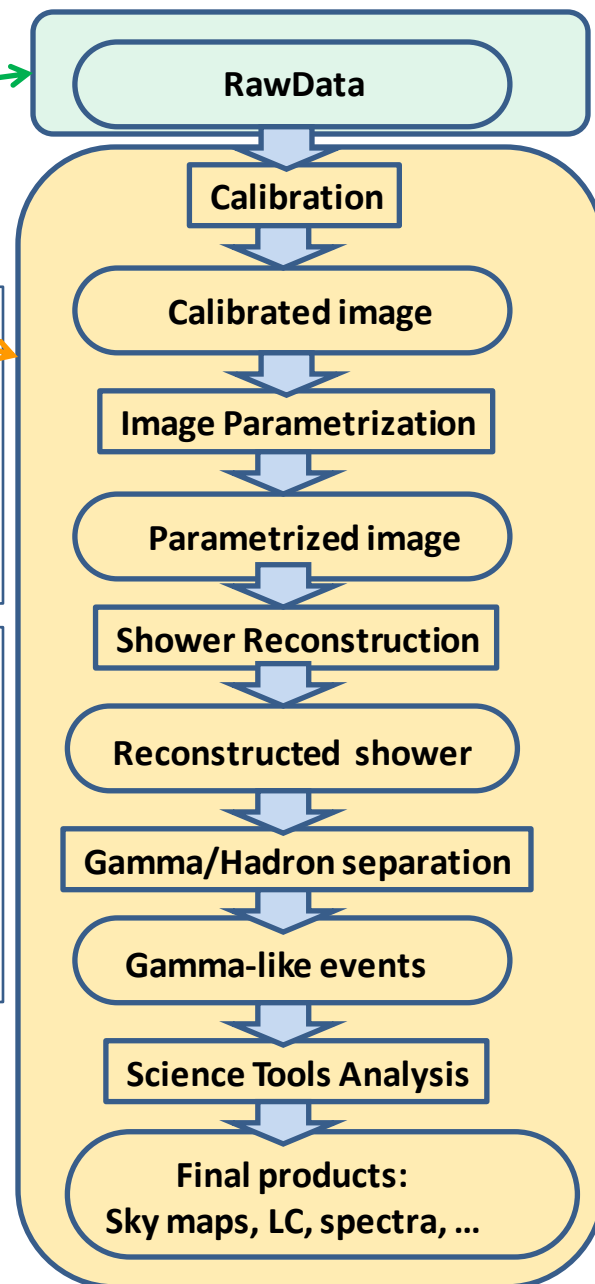
- sim_telarray sw [Bernlohr 2008, AIP Conf. Proc., Vol. 1085, 874-877]

Analysis

- No official CTA analysis software is available
- Analysis sw coming from the current IACTs:
- more than one analysis chain for each experiment (HESS, MAGIC, VERITAS)
- some of them are proprietary software

Computing needs for a typical MC production

- ~150k jobs
- corsika + sim_telarray in a single job (about ~3h)
- producing **about 200 TB**

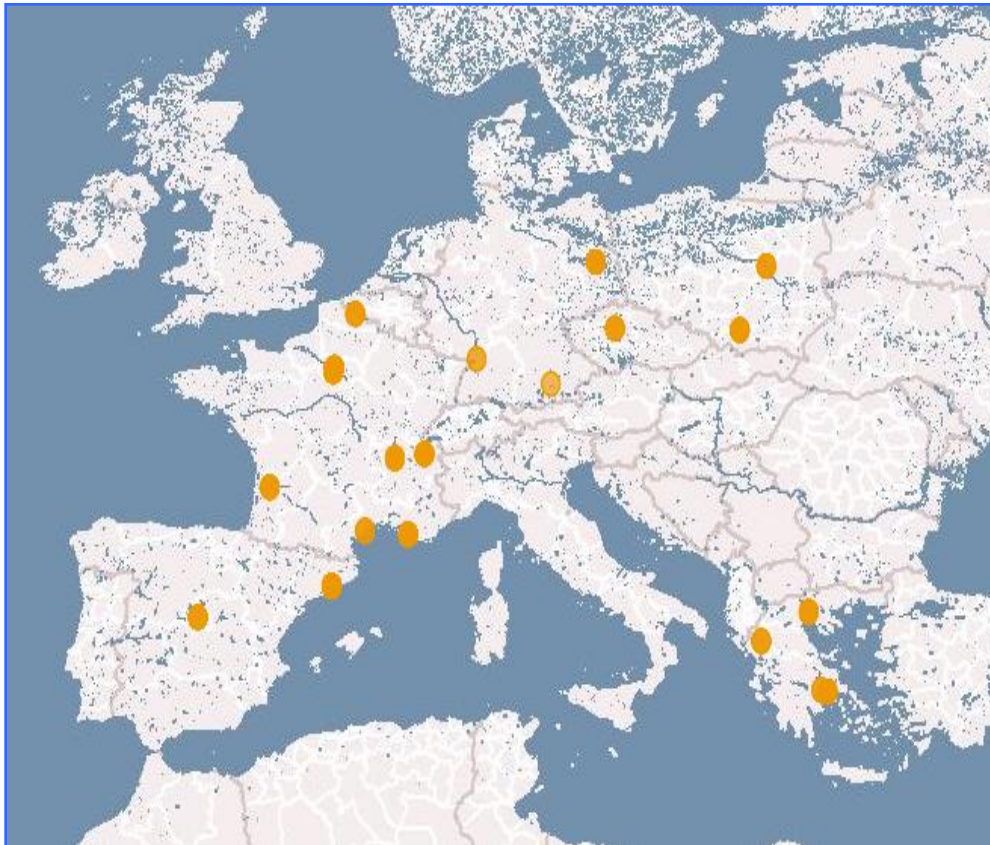


MC simulation on the EGI Grid: past experience

Grid used for MC productions since 2008

- Submission and configuration tool based on GANGA
- Parallel productions on local clusters

17 EGI Sites spread in 6 countries



Computing resources

1000 – 2000 CPUs simultaneously available based on past experience

Storage resources

- **390 TB** distributed in all Sites and shared with other VOs
- **430 TB dedicated to CTA:**
 - 1) IN2P3-CC Disk and Tape
 - 2) DESY-ZEUTHEN
 - 3) CYFRONET

MC simulations on the Grid: DIRAC

- First proposal of using DIRAC to handle MC simulations (production and analysis) one year ago by LUPM and UB teams

Goals:

- Optimise MC production
- Provide tools for User Analysis

Motivations for using DIRAC:

- Integrate *heterogeneous resources*
- Implementation of the *Pilot Job mechanism* -> *high efficiency*
- Central management of *VO policies*
- *Flexible framework* allowing an easy customization
- Handle bulk submission
- Fine grained job monitoring/accounting (web based)
- *Save MetaData* informations in the DIRAC File Catalog allowing users' queries through Metadata
- *Fault-tolerance* design for core services
- User interface (python API, command-line, web portal)

Current use of DIRAC

What's done so far?

DIRAC instance for CTA

- 3 servers at PIC:
 - Central services
 - Agents
 - Sandboxes
- 1 server at CC-IN2P3:
 - Storage Element Service for log files and Failover
 - Configuration System mirror

Current use of DIRAC

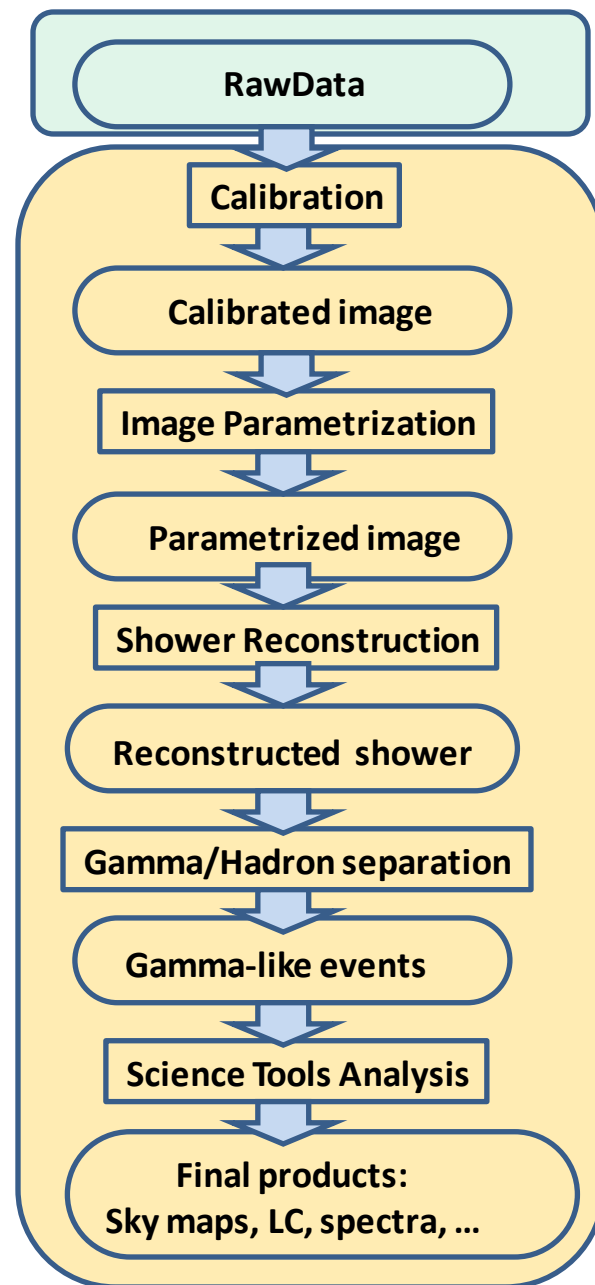
What's done so far?

MC Production

- Implementation of DIRAC modules to run massive simulations (corsika+sim_telarray)
- **Use of the DIRAC File Catalog** to save Metadata (identify specific functionalities to handle CTA Data Model)
- Not tested yet at large scale, since the next MC campaign is not yet ready

MC Analysis

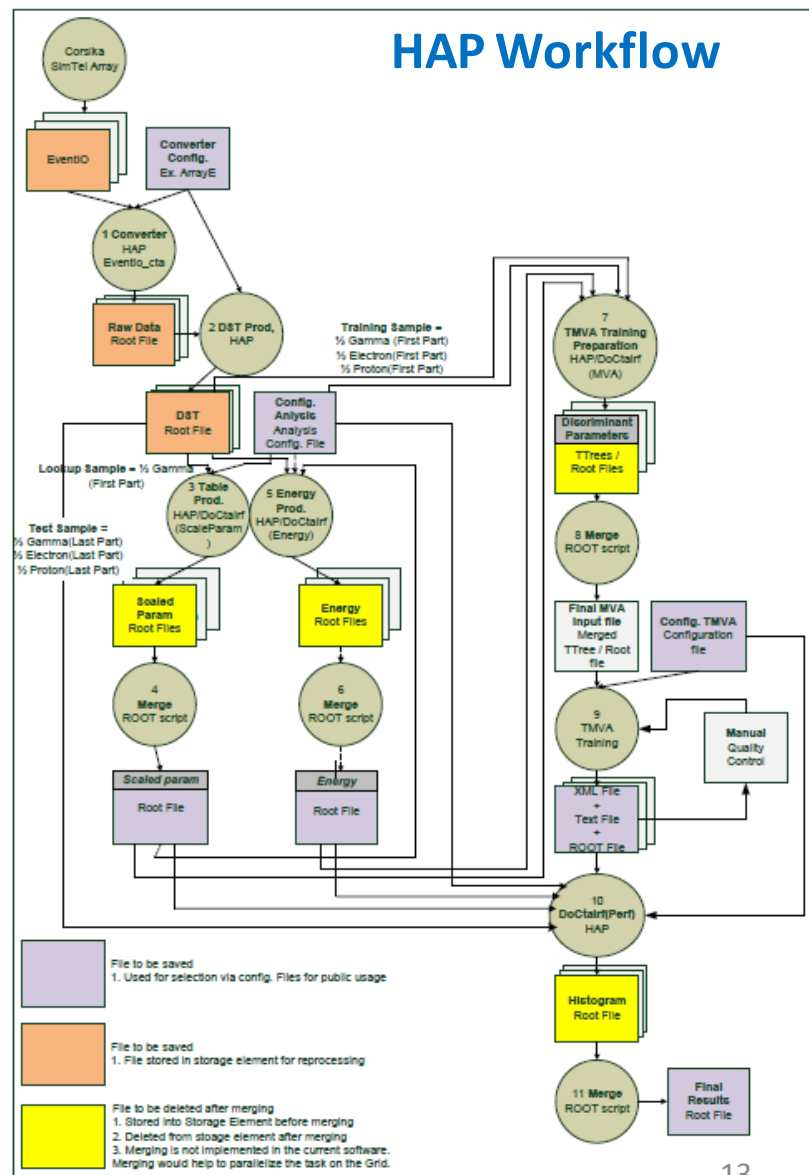
- Implementation of **DIRAC modules** to run some of the analysis chains used in CTA:
 - 1) **read_hess, HAP (from HESS)**
 - 2) Recently some work has started for the **MAGIC sw**
 - 3) **High-level analysis sw**
- Work on **end-user documentation**
- Got first user feedback



Current use of DIRAC: HAP analysis chain

Complex workflow

- Many interdependent steps linked by I/O
 - Different CPU and I/O profiles
 - Different data products (temporary and persistent files)
 - Quality checks on data products
 - Auxiliary jobs: Merging/ Format Conversion
-
- Currently: a DIRAC module for each step
 - Next: combine them into workflows



http://www.cta-observatory.org/ctawpcwiki/index.php/CEIN/DIRAC_evaluation_for_CTA

(CTA internal)

Documentation for users
with no knowledge about
the Grid

Section dedicated
to examples for CTA
applications



- 1 Introduction
- 2 Support
- 3 How to install the DIRAC client
 - 3.1 Prerequisites
 - 3.2 Getting your Grid identity
 - 3.2.1 Downloading your certificate from the browser
 - 3.2.2 Converting your certificate from P12 to PEM format
 - 3.3 Installation
 - 3.4 Configuration
 - 3.5 Using your client
 - 3.6 Updating your client installation
- 4 How to use the DIRAC client
 - 4.1 Introduction
 - 4.2 Data Management concepts
 - 4.2.1 Browse the File Catalog
 - 4.2.2 Retrieve files from the grid storage
 - 4.2.3 Upload a file to the grid storage and register it in the File Catalog
 - 4.2.4 Remove a file from the grid storage and the File Catalog
 - 4.3 Job Submission using the DIRAC API
 - 4.3.1 Input Data
 - 4.3.2 Output Data
 - 4.3.3 How to use Sandboxes
 - 4.3.3.1 InputSandbox (ISB)
 - 4.3.3.2 OutputSandbox (OSB)
 - 4.3.4 Job executable
 - 4.3.5 CPU Time
 - 4.3.6 Job Group
 - 4.3.7 Monitor your jobs and retrieve the output
 - 4.3.8 Select your jobs according to given criteria
 - 4.4 Job submission through the DIRAC web portal
- 5 How to run CTA applications using the DIRAC client
 - 5.1 Prepare your production
 - 5.2 How to run corsika
 - 5.3 How to run sim_telarray
 - 5.3.1 Description
 - 5.4 How to run read_hess
 - 5.4.1 How to run read_hess through the DIRAC web portal
 - 5.5 How to run HAP
 - 5.5.1 Rawdata production
 - 5.5.2 DST production
 - 5.5.3 Rawdata and DST production in a single job
 - 5.5.4 Lookup tables production
 - 5.5.5 TMVA training preparation
 - 5.5.6 Speedup the job submission for massive productions
 - 5.5.6.1 Rawdata production
 - 5.5.6.2 DST production
 - 5.5.6.3 Rawdata and DST production in a single job
- 6 Links and talks related to this topic

User experience (I)

(by P. Jean/IRAP)

Scientific goal:

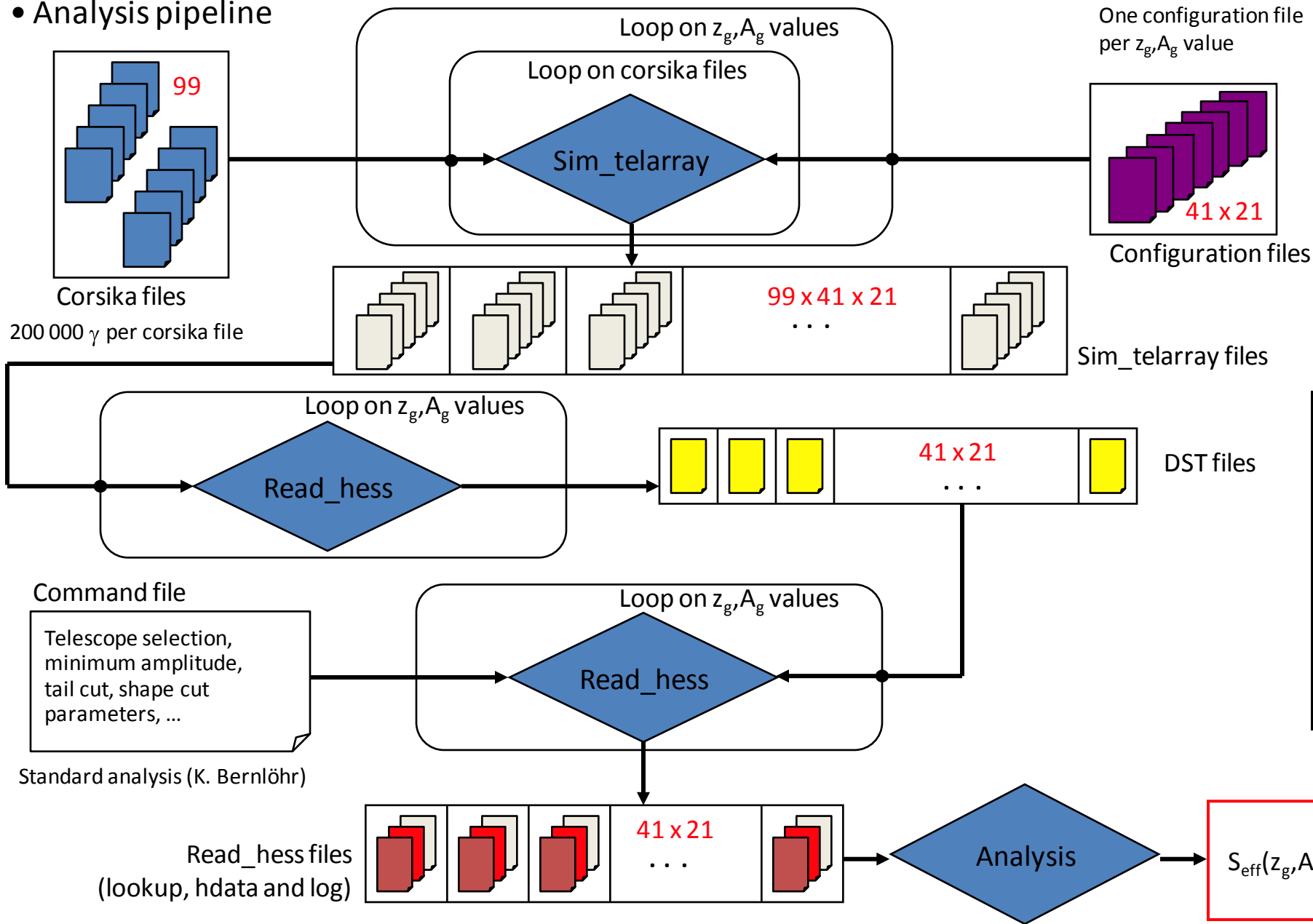
Calculate CTA effective area as a function of source angular position with respect to global CTA axis. The goal is to determine if we can gain effective area and global field of view if CTA is in a 'divergent mode' compared to the 'normal mode' for which all telescope point at the same direction

Program:

- Source scan over (21x21) bins in (zenith, azimuth) made twice for 'normal' and 'divergent' modes
- *High computing needs*
- *Use DIRAC to access Grid resources*

User experience (II)

Analysis pipeline

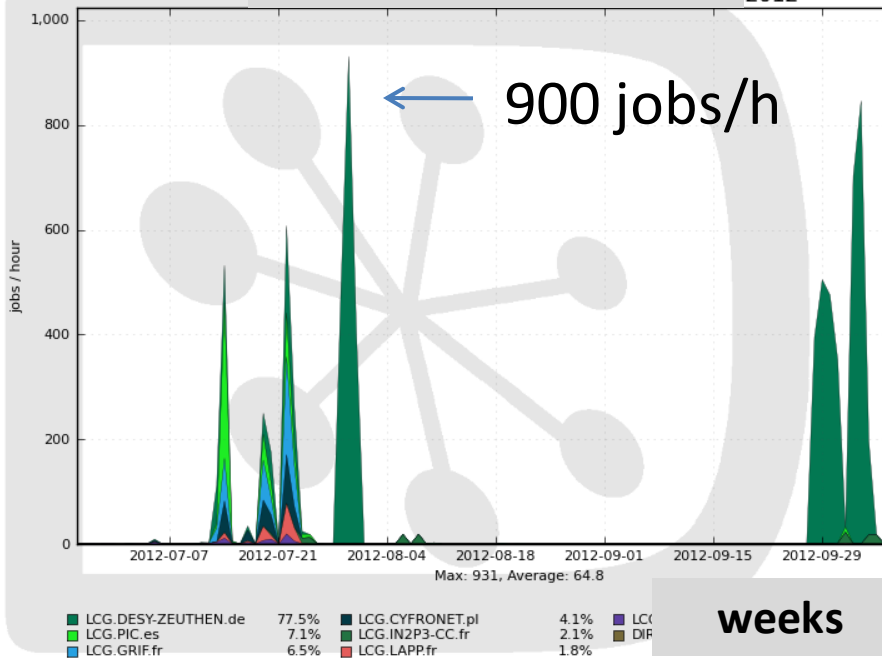


User experience (III)

About 4 weeks to run the program (July-October 2012):

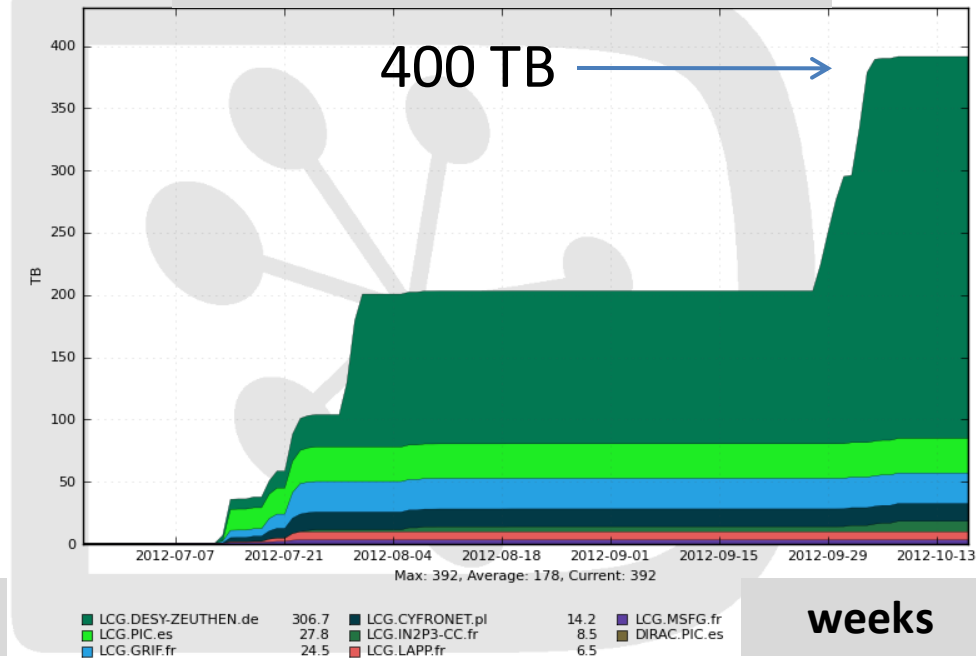
- 180k jobs
- 400 TB processed
- 12 TB Output Data

Job rate by Site



Generated on 2012-10-1

Cumulative Input Data by Site



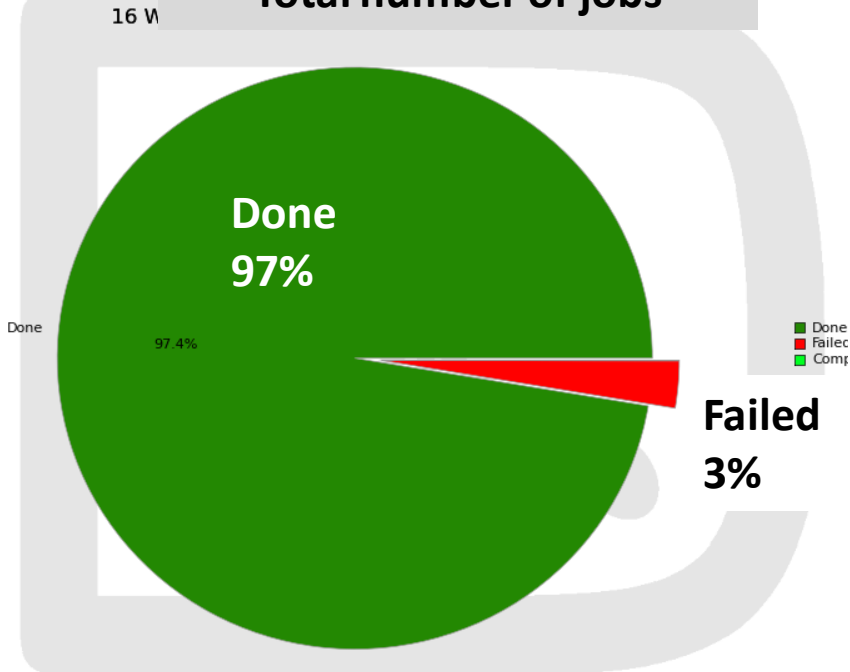
Generated on 2012-10-18 14:27:07 UTC

User experience (IV)

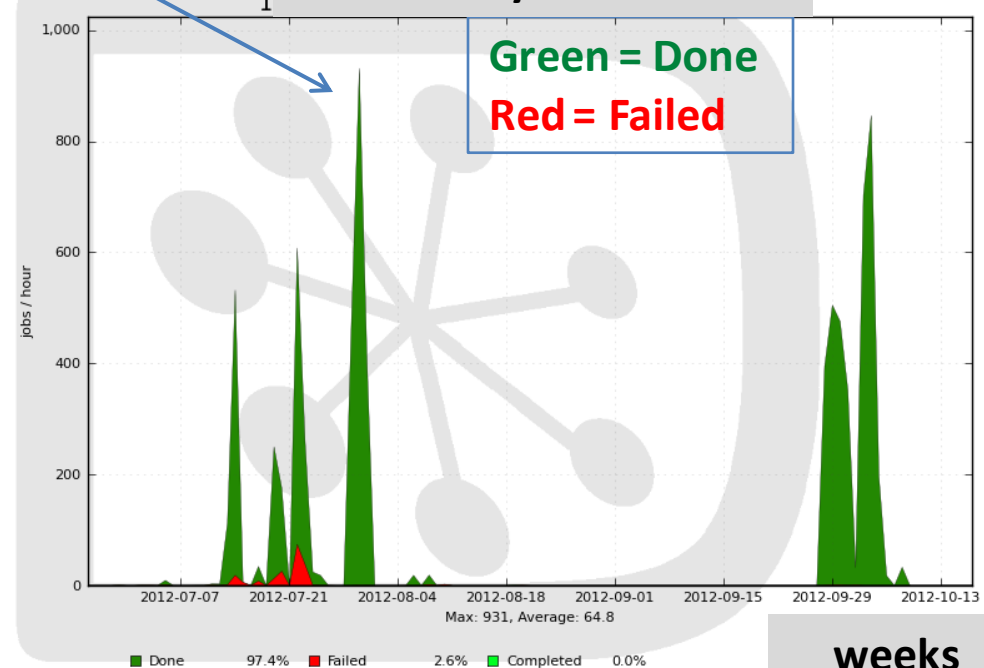
Job Success rate:

- overall > **97%** (including first phase of test)
- 100% in last bunches

Total number of jobs



Job rate by Final Status

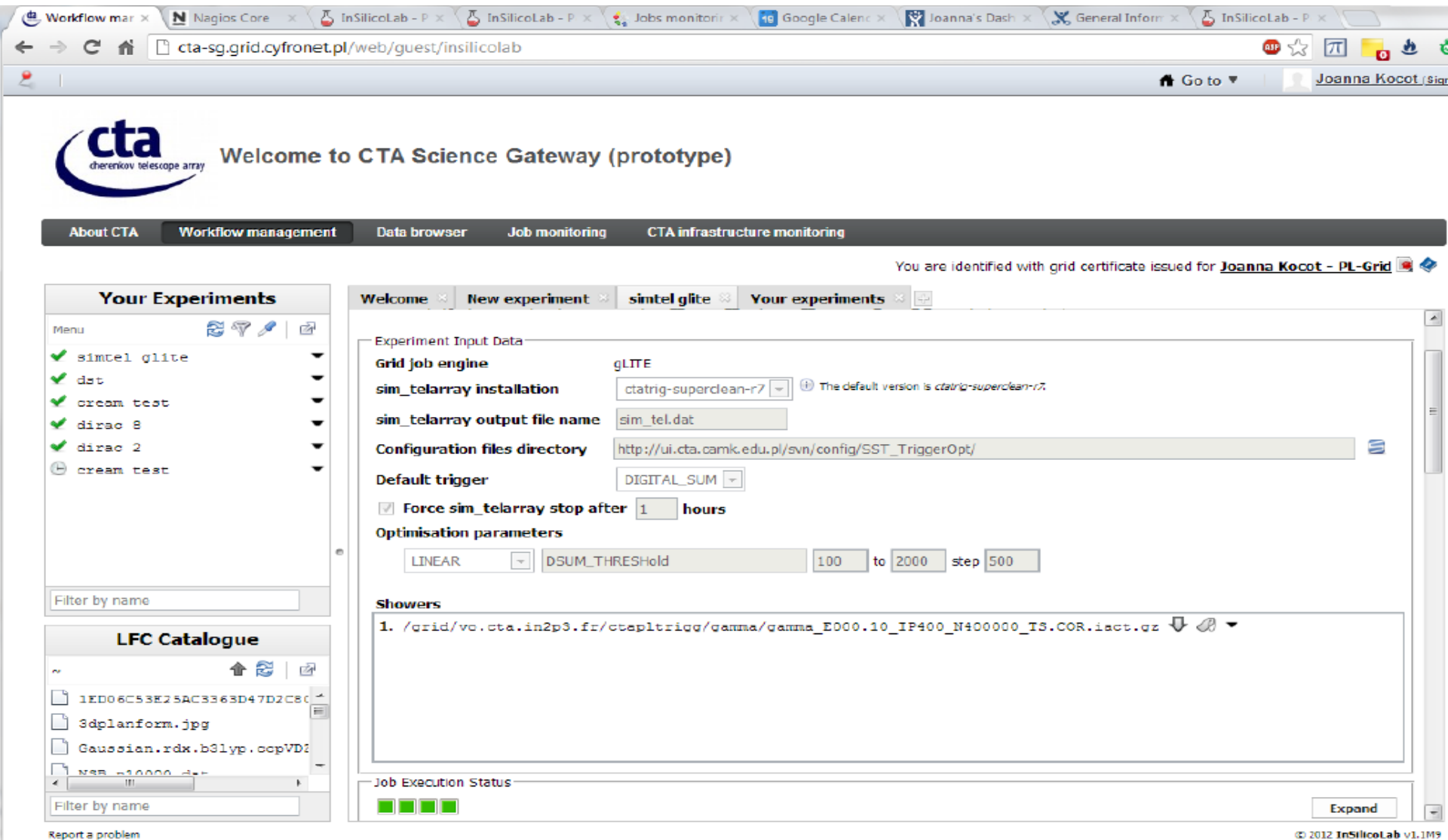


weeks

Generated on 2

Toward a CTA Science Gateway

Work in progress to integrate DIRAC CTA modules in a web portal (prototype by CYFRONET)



The screenshot shows a web browser window displaying the CTA Science Gateway (prototype) interface. The browser's address bar shows the URL `cta-sg.grid.cyfronet.pl/web/guest/insilicolab`. The page features a navigation bar with links: About CTA, Workflow management, Data browser, Job monitoring, and CTA infrastructure monitoring. The main content area is titled "Your Experiments" and includes a sidebar with a "Menu" and a "LFC Catalogue". The "Your Experiments" section contains a list of experiments: `simtel glite`, `dat`, `cream test`, `dirac 8`, `dirac 2`, and `cream test`. The "LFC Catalogue" section shows a list of files: `1ED06C53E25AC3363D47D2C8C`, `3dplanform.jpg`, `Gaussian.rdx.b31yp.copVD`, and `N3B_010000.dat`. The main content area displays the "Experiment Input Data" form for the `simtel glite` experiment. The form includes fields for "Grid job engine" (glite), "sim_telarray installation" (ctatrig-superclean-r7), "sim_telarray output file name" (sim_tel.dat), "Configuration files directory" (http://ui.cta.camk.edu.pl/svn/config/SST_TriggerOpt/), "Default trigger" (DIGITAL_SUM), and "Optimisation parameters" (LINEAR, DSUM_THRESHOLD, 100 to 2000 step 500). The "Showers" section lists a single shower: `1. /grid/vc.cta.in2p3.fr/ctapltrigg/gamma/gamma_E000.10_IP400_N400000_IS.COR.iact.gz`. The "Job Execution Status" section shows four green squares, indicating successful execution. The user is identified as `Joanna Kocot - PL-Grid`.

Conclusions and perspectives

- CTA is the next generation instrument in the VHE gamma-ray astronomy
- It will be producing **several PB/year**
- **DIRAC is currently used to handle MC simulations on the Grid**
- We found out that all needed functionalities for MC production are covered
- User Analysis experience is also positive
- **DIRAC File Catalog** is currently being tested
- Work in progress to integrate DIRAC CTA modules in a **CTA Science Gateway prototype**
- Need to define a Computing Model for future CTA operation phase