

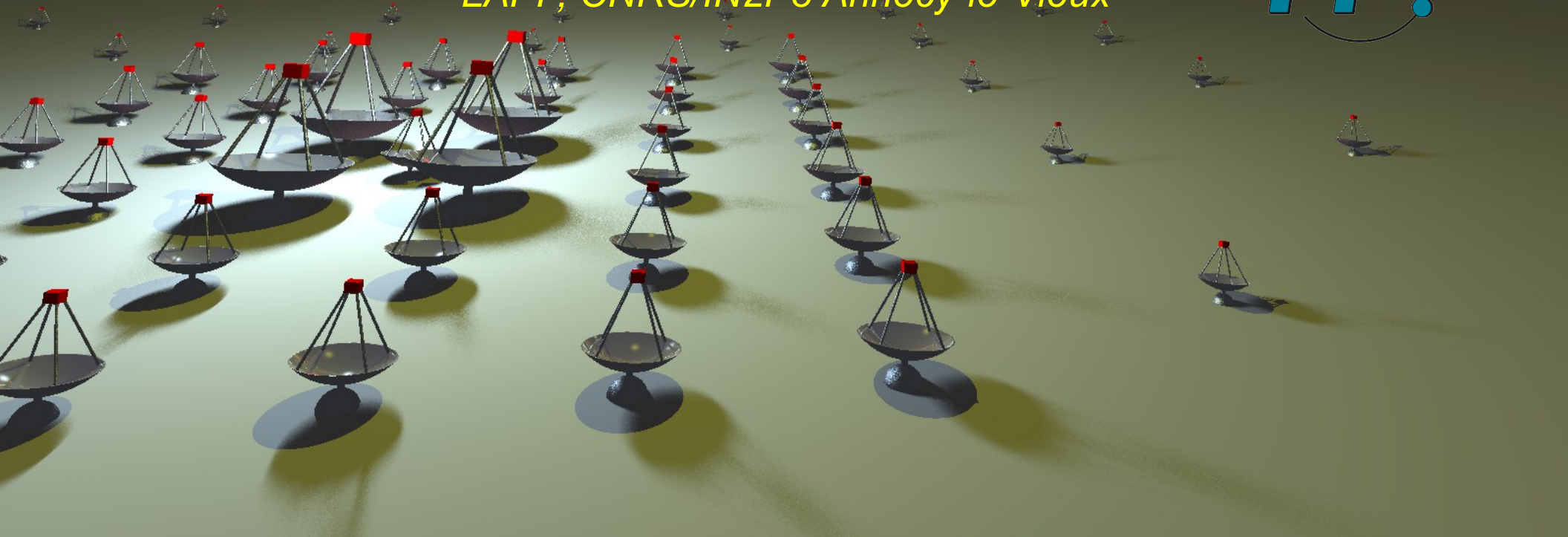
The CTA Computing Grid Project



Nukri Komin



LAPP, CNRS/IN2P3 Annecy-le-Vieux



Outline

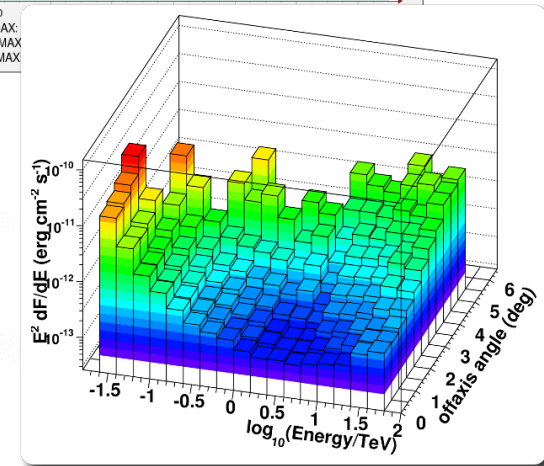
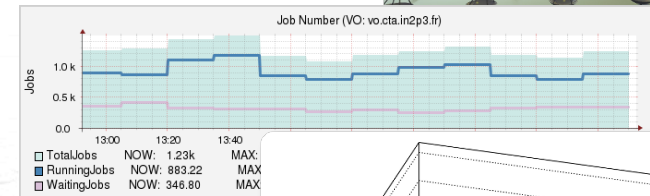
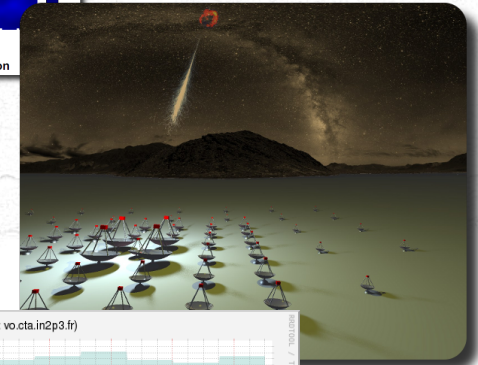
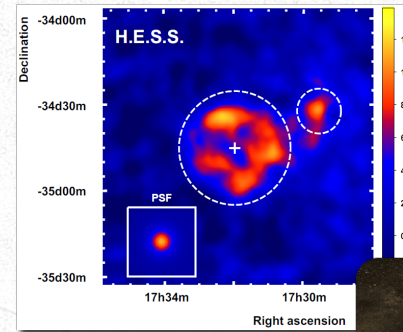
➔ Cherenkov Telescope Array

- next generation observatory for gamma-ray astronomy

➔ Computing Grid

- Monte Carlo simulations
- developments
- results

➔ Future data pipe line and Science Analysis System

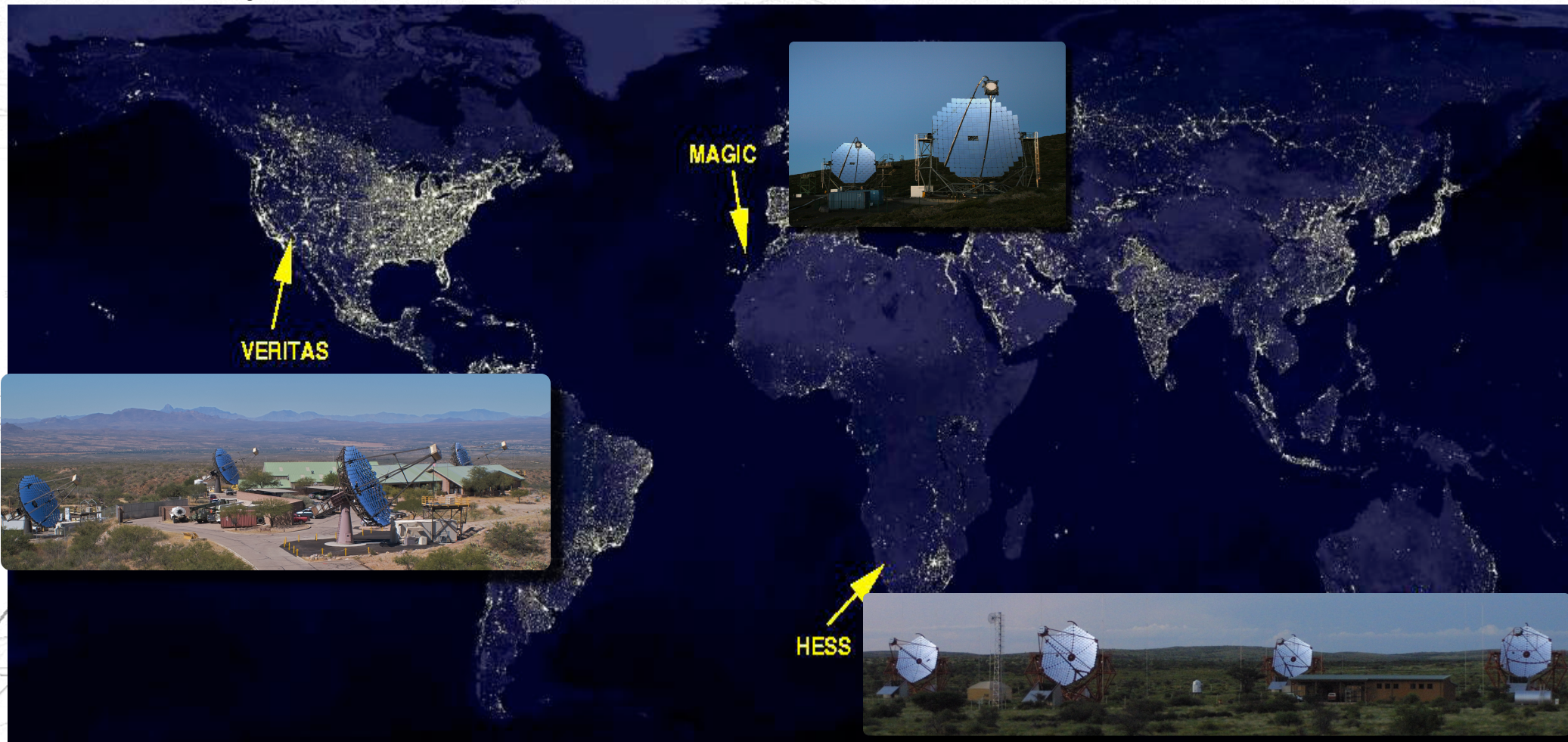


Gamma-Ray Astronomy

⇒ astronomy at TeV (10^{12} eV) energy scale

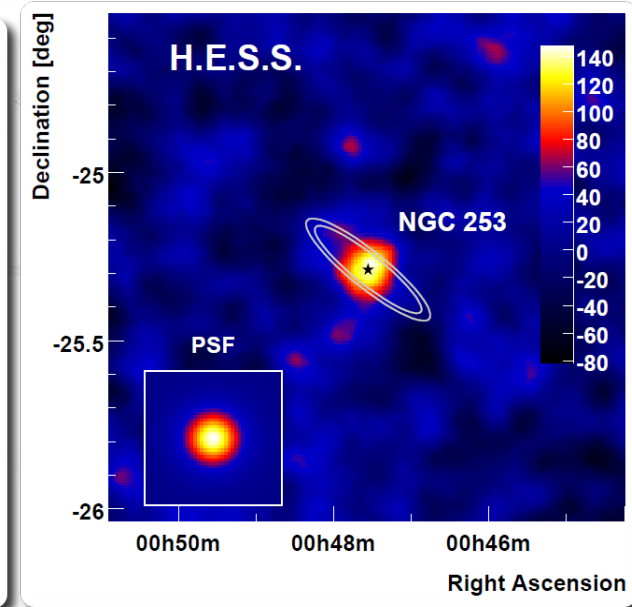
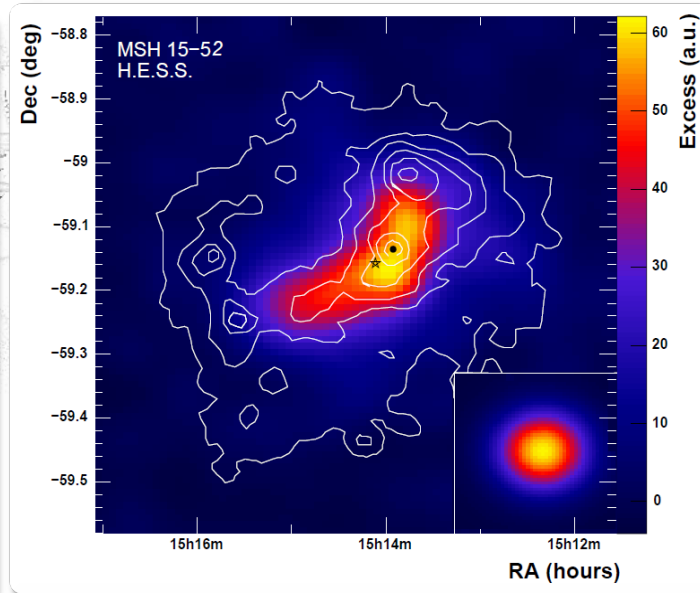
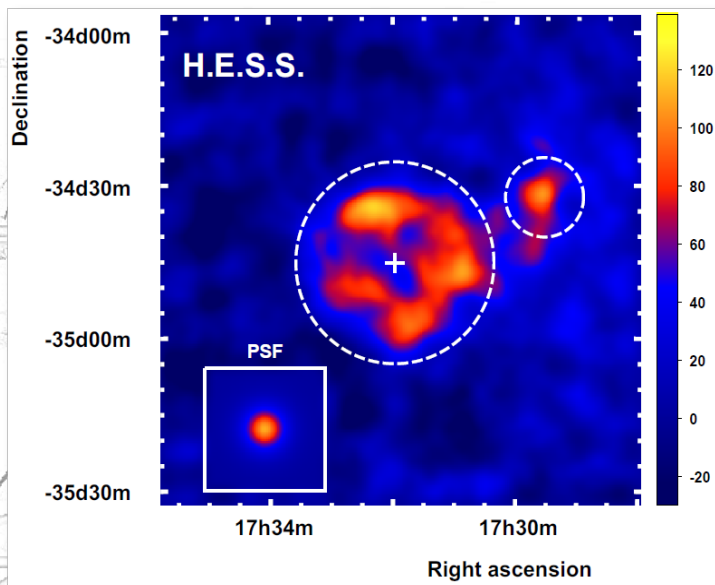
- billion times energy of X-rays

⇒ currently 3 experiments with 2 – 5 telescopes

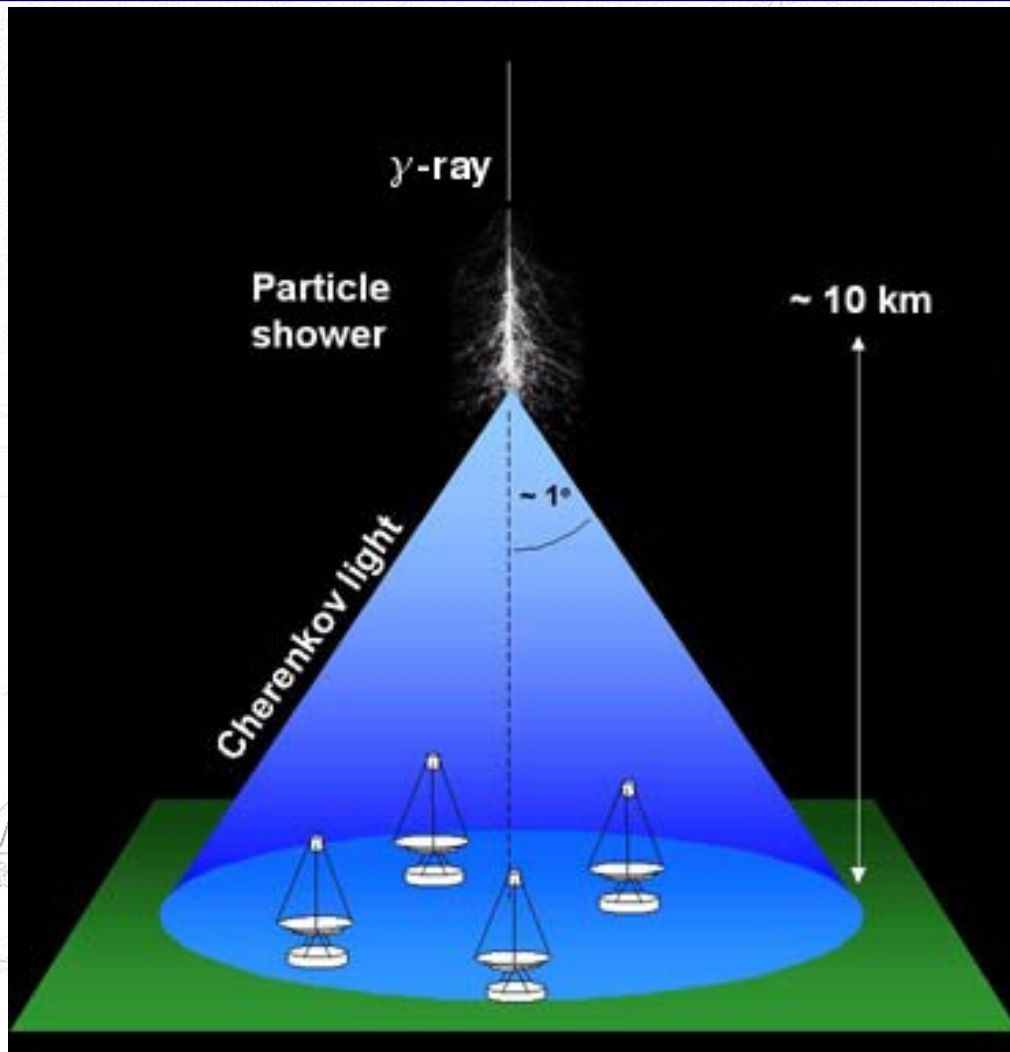


Gamma-Ray Astronomy

- 1st discovery in 1989
- current instruments 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



Imaging Air Cherenkov Technique



- ⇒ photon creates air shower of tens of thousands of secondary particles
- ⇒ Cherenkov emission of shower particles
- ⇒ detection with multiple optical telescopes
- ⇒ total area
→ sensitivity
- ⇒ mirror surface
→ sensitivity, energy threshold
- ⇒ number of telescopes
→ angular resolution and background rejection

Next Generation: CTA

⇒ Cherenkov Telescope Array

- consortium of >800 scientists in 25 countries
- aim: 10x sensitivity of current instruments
- 2 sites: south and north
- operate as observatory

⇒ core of large telescopes:

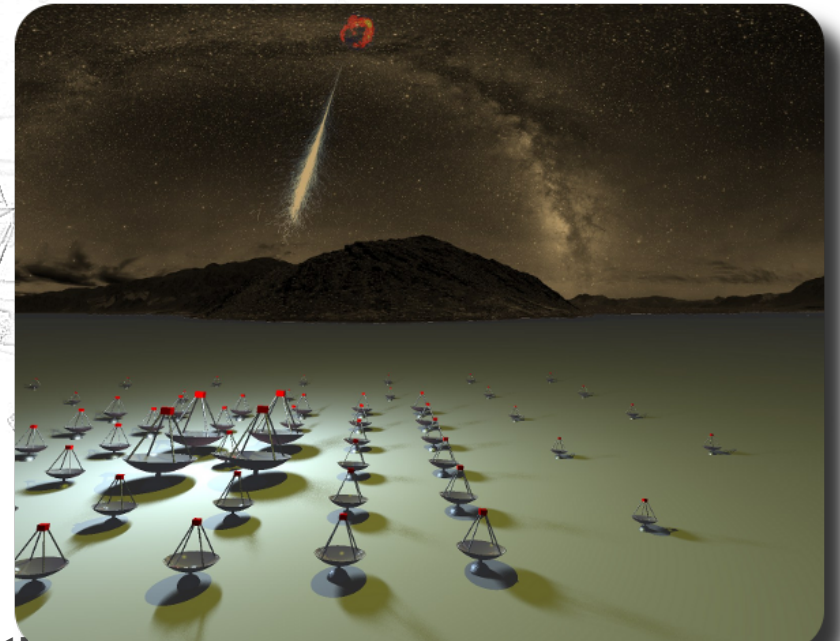
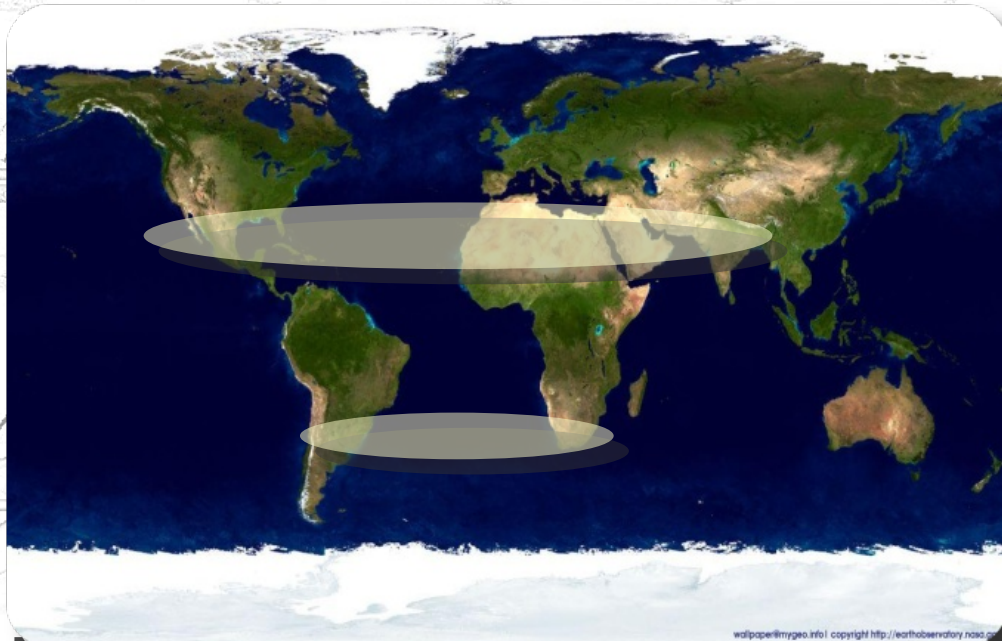
- high sensitivity at low energy

⇒ medium size telescopes

- energy band: 100 GeV – 10 TeV

⇒ small telescopes

- large area for multi-TeV



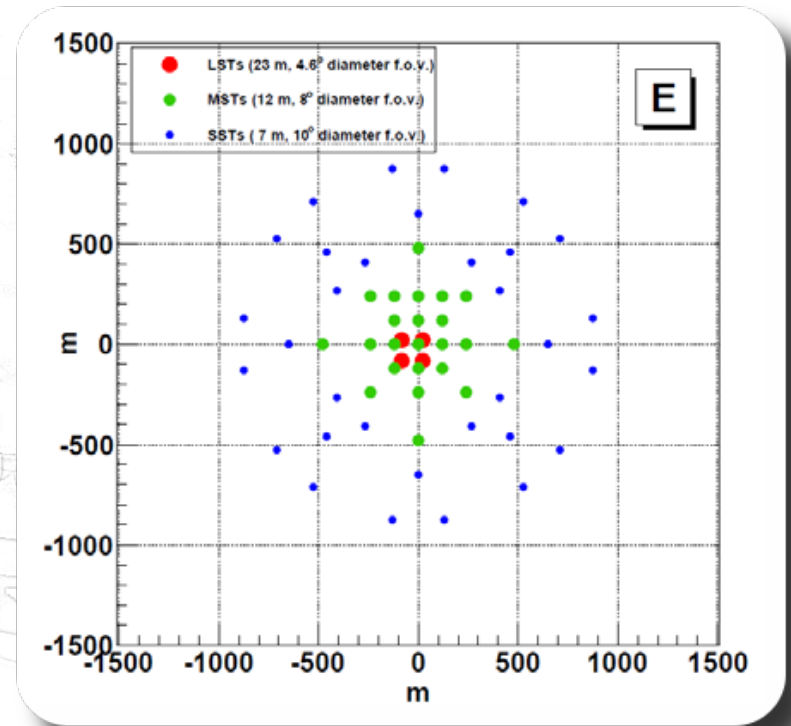
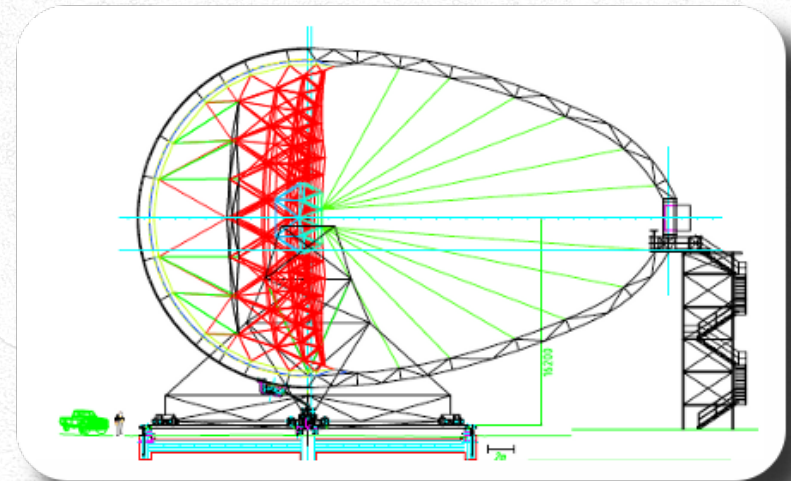
CTA: Preparatory Phase

⇒ 2010 – 2013

⇒ accepted project in FP7

⇒ address questions of construction and operation of large telescope array

- find sites
- build prototypes of telescopes
- define optimal array layout
 - testing sub-arrays out of hypothetical mega-array (295 telescopes)
- prepare data pipeline and data distribution



CTA Computing and Grid Approach

⇒ Preparatory Phase

- Monte Carlo simulations
- air showers of 10^4 particles
- simulation of mega-array
- analysis of many layouts and telescope designs

⇒ operation

- raw data on remote site
- data calibration and particle reconstruction
- analysis and data publication

⇒ no dedicated CTA computing centres yet

⇒ computing done in participating institutes individually

⇒ idea: join the computing via Grid technology

⇒ give user access to all data and combined computing power

CTA Virtual Organisation

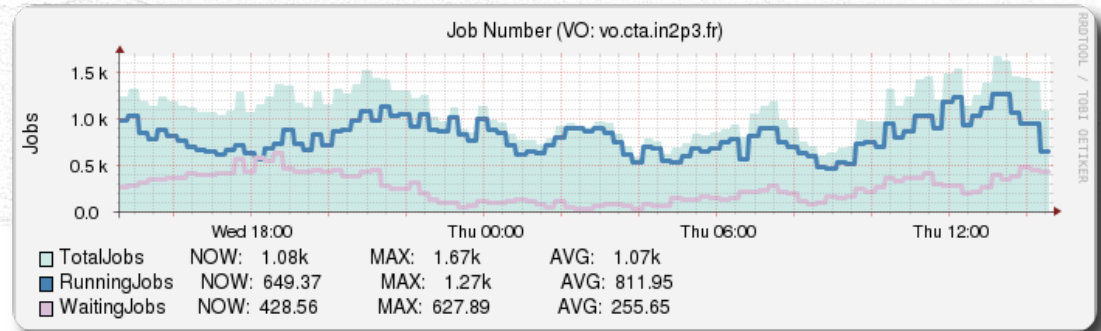
➔ virtual organisation: vo.cta.in2p3.fr

- VOMS and LFC server hosted at CC Lyon, dedicated WMS at LAPP

➔ 14 sites in 6 countries: 37k CPUs shared with other VOs

- 1000 – 2000 CPUs used by CTA simultaneously

➔ ~500 TB of storage distributed over sites



Country	City	Name	CPU		storage
			Logical	SI2000	size [GB]
France	Paris	GRIF	7789	16273608	112827
	Lyon	IN2P3-CC	9068	19459928	16106 +several 100TB on tape
	Annecy	IN2P3-LAPP	816	1952688	5499
	Bordeaux	M3PEC	432	907200	11306
	Montpellier	MSFG-OPEN	152	300200	5326
	Meudon	OBSPM	112	401184	11646
Germany	Zeuthen	DESY-ZN	768	2602752	90000
	Munich	MPPMU	2352	6644400	0
Poland	Warsaw	CAMK	0	0	4921
	Cracow	CYFRONET-LCG2	8712	18748224	73101
Spain	Madrid	CIEMAT-LCG2	944	2765920	231
	Barcelona	PIC	3216	8728224	11012
Czech Republic	Prague	praguelcg2	2956	6104140	0
Greece	Thessaloniki	GR-01-AUTH	230	484150	2953

Monte Carlo Simulation on the Grid

➔ air showers and Cherenkov light

- CORSIKA [<http://www-ik.fzk.de/corsika/>]
- high memory requirement (up to 4GB RAM for proton showers)
- intermediate file: 1–2 GB (for 10000 gammas or 1000 protons)

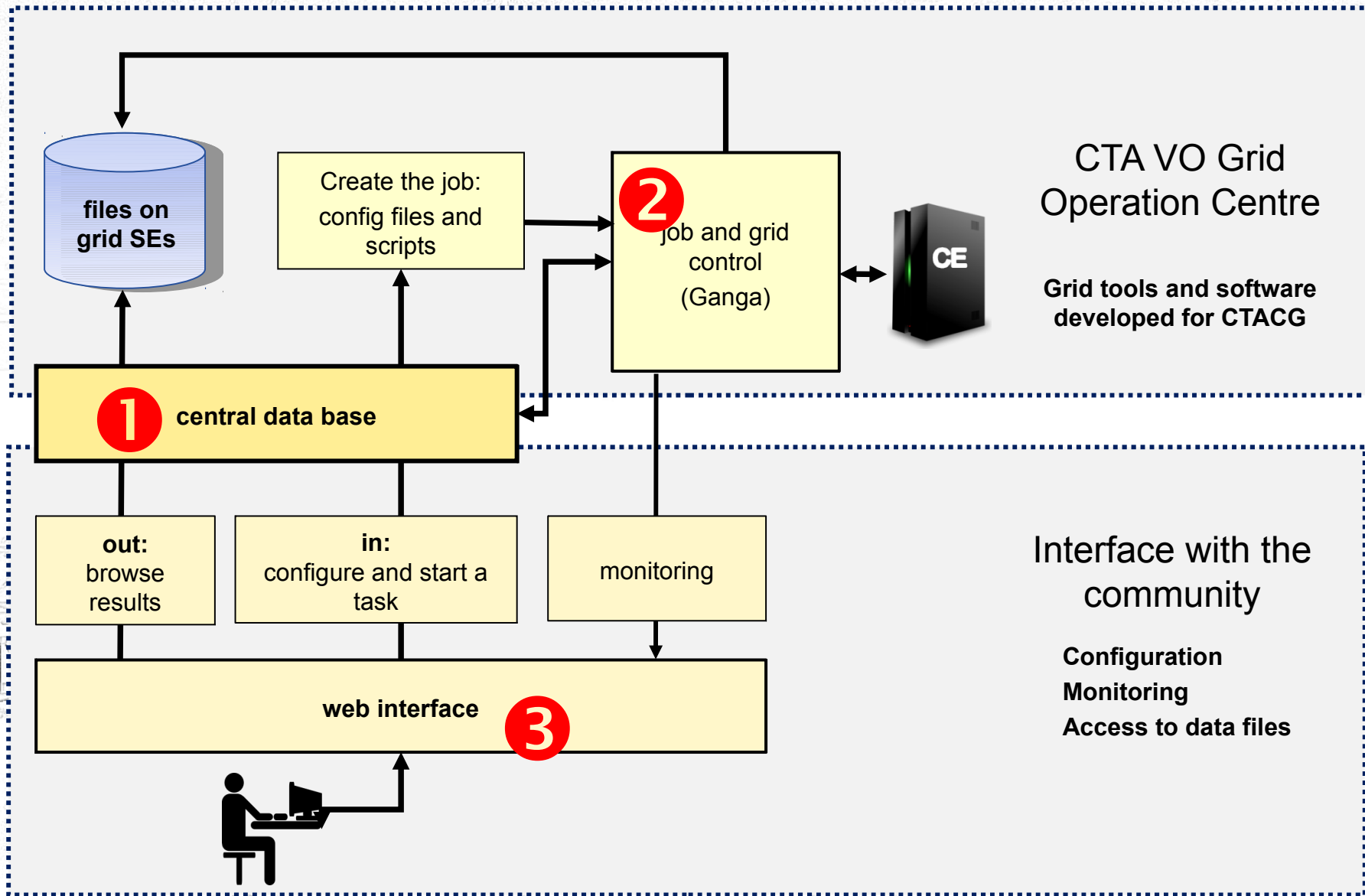
➔ simulation of telescope response

- sim_telarray [Bernlöhr 2008, AIP Conf. Proc., Vol. 1085, 874-877]
- 295 telescopes
- final file size: 100 – 400 MB
- current production: 100000 files

➔ corsika + sim_telarray in 1 job: ~3h

➔ current production: save some corsika files on large SEs (Lyon, DESY Zeuthen, Cyfronet) to build up a corsika stock

Automated Job Submission



EasiJob



➔ **Easy Integrated Job Submission**

➔ tool developed at LAPP within the MUST framework

- MUST – computing centre for LAPP and Savoie University, Tier 2

➔ configuration of task

➔ configuration of site classes

➔ checkout and book-keeping

➔ job submission

- EasiJob commands are stored in data base
- job submission and monitoring using Ganga
- loop running in Ganga checks for new commands
- job submission pauses on maximum number of waiting jobs

Configuration

- ➔ configuration of parameter set with key words and default values
- ➔ input sandbox template where #keyword will be replaced with value during job submission
- ➔ set of parameter values in data base for each task

The screenshot displays the EasiJob configuration interface. On the left, a tree view shows the configuration structure under 'Parameters'. The selected configuration is '20 - INPUTS_CTA_ULTRA3_gamma'. The tree includes folders for 'Generic shower parameters', 'Random number generator', and 'Primary particle options'. Under 'Primary particle options', there are sub-folders for 'PRMPAR', 'ESLOPE', 'ERANGE', 'THETAP', and 'VIEWCONE', each containing specific parameters like 'ParticleId0', 'EnergySlope', 'EnergyMin', 'EnergyMax', 'ThetaP0', 'ThetaP1', 'PhiP0', and 'PhiP1'.

The right panel shows the configuration details for the selected parameter 'ParticleId0':

- Parameter: ParticleId0
- ParamIdentifier: #ParticleId0#
- Default value: 1
- Parameter type: basic type
- Parameter basic type: int
- Parameter unit: no unit

Buttons for 'Save' and 'Save all parameters to DB' are visible at the bottom of the configuration panel.

Site Classes

→ sites are added to class

→ JDL requirements set individually

The screenshot shows the EasiJob web interface in a Firefox browser. The URL is <http://lappw...www/TaskReq/>. The interface includes a sidebar with the EasiJob logo and the lapp logo (Annee-les-Vieux). The main content area displays the configuration for requirements defined in the database.

Requirements defined in DB :

Select a requirement : duplicate as

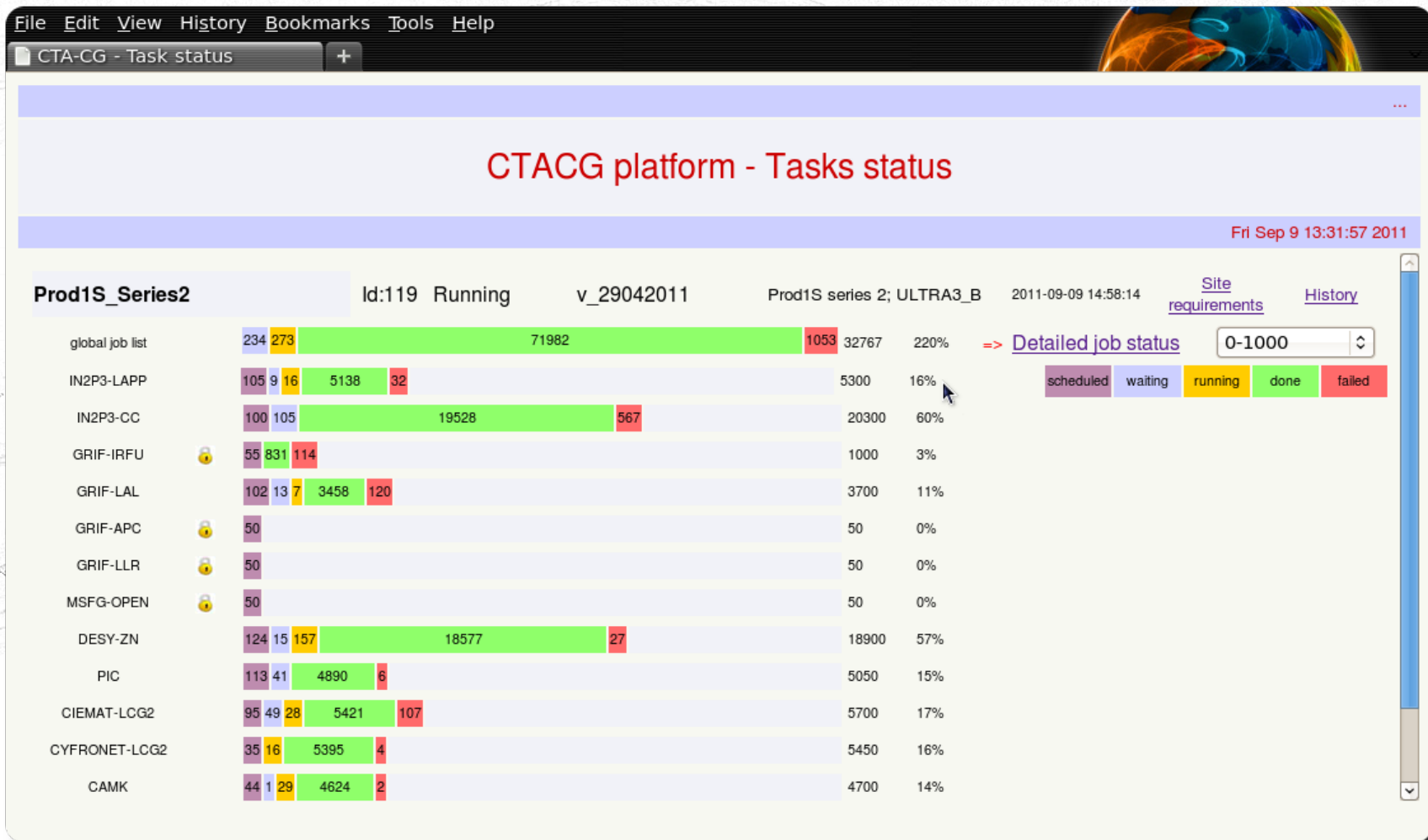
Selected requirements : CTAProd_RAM_2GB

General requirement :

Site name	Status	Max/Min CPU (Mo)	Max/Min RAM (Mo)	other requirement
IN2P3-LAPP (1)	enabled	no ▾ 0	no ▾ 0	none
IN2P3-CC (2)	enabled	>= ▾ 2000	no ▾ 0	none
GRIF-IRFU (3)	enabled	no ▾ 0	no ▾ 0	none
GRIF-LAL (4)	enabled	no ▾ 0	no ▾ 0	none
GRIF-APC (5)	enabled	no ▾ 0	no ▾ 0	none
GRIF-LLR (6)	enabled	no ▾ 0	no ▾ 0	none
GRIF-LPNHE (7)	enabled	no ▾ 0	no ▾ 0	none
MSFG-OPEN (8)	enabled	no ▾ 0	no ▾ 0	none
MPPMU (9)	enabled	>= ▾ 2000	>= ▾ 2000	none
DESY-ZN (10)	enabled	>= ▾ 2000	>= ▾ 2000	none
UNI-DORTMUND (11)	enabled	>= ▾ 2000	>= ▾ 2000	none
PIC (12)	enabled	>= ▾ 2000	no ▾ 0	none
CIEMAT-LCG2 (13)	enabled	no ▾ 0	no ▾ 0	none
CYFRONET-LCG2 (14)	enabled	>= ▾ 2000	>= ▾ 2000	none
CAMK (15)	enabled	no ▾ 0	no ▾ 0	none
GR-01-AUTH (16)	enabled	no ▾ 0	no ▾ 0	none
PRAGUE-LCG2 (17)	enabled	no ▾ 0	no ▾ 0	none

Submission and Monitoring

- ⇒ submission of requested jobs
- ⇒ limit for *waiting* jobs, the rest remains *scheduled*
- ⇒ re-submission of failed jobs



Bookkeeping

➔ check if file is registered in LFC

➔ production parameters and file location (logical and physical) stored in data base

Firefox CTACG platform

CTACG platform - job status

Wed Sep 14 18:06:15 2011

Job number	Run	Status	LFN	Grid site name	Time stamp	Ganga job number
2	50001	done	<code>/grid/vo.cta.in2p3.fr/Simulation/corsika/log/run050001_dir.tar.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/Prod1S/gamma/sim_hessarray/cta-prod1/0.0deg/Data/gamma_20deg_90deg_run50001__cta-prod1_desert.simhess.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/sim_hessarray/cta-prod1/0.0deg/Histograms/gamma_20deg_90deg_run50001__cta-prod1_desert.hdata.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/sim_hessarray/cta-prod1/0.0deg/Log/gamma_20deg_90deg_run50001__cta-prod1_desert.log.gz</code>	IN2P3-LAPP	Wed Aug 3 17:10:24 2011	16806
3	50002	done	<code>/grid/vo.cta.in2p3.fr/Simulation/corsika/log/run050002_dir.tar.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/Prod1S/gamma/sim_hessarray/cta-prod1/0.0deg/Data/gamma_20deg_90deg_run50002__cta-prod1_desert.simhess.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/sim_hessarray/cta-prod1/0.0deg/Histograms/gamma_20deg_90deg_run50002__cta-prod1_desert.hdata.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/sim_hessarray/cta-prod1/0.0deg/Log/gamma_20deg_90deg_run50002__cta-prod1_desert.log.gz</code>	IN2P3-LAPP	Wed Aug 3 16:32:36 2011	16807
4	50003	done	<code>/grid/vo.cta.in2p3.fr/Simulation/corsika/log/run050003_dir.tar.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/Prod1S/gamma/sim_hessarray/cta-prod1/0.0deg/Data/gamma_20deg_90deg_run50003__cta-prod1_desert.simhess.gz</code> <code>/grid/vo.cta.in2p3.fr/Simulation/sim_hessarray/cta-prod1/0.0deg/Histograms/gamma_20deg_90deg_run50003__cta-prod1_desert.hdata.gz</code>	IN2P3-LAPP	Wed Aug 3 16:32:37 2011	16808

User's Analyses

⇒ currently: many users download simulation files for local analysis

⇒ goal: bring user to the Grid

- need user interface with easy installation and easy use
- need analysis software

⇒ testing Dirac



- server running in Barcelona (PIC)
- preparation of analysis scripts
- currently ROOT, next step: H.E.S.S. and Magic analysis software

⇒ currently no dedicated CTA software

- use of proprietary software of individual instruments
- not shared among CTA

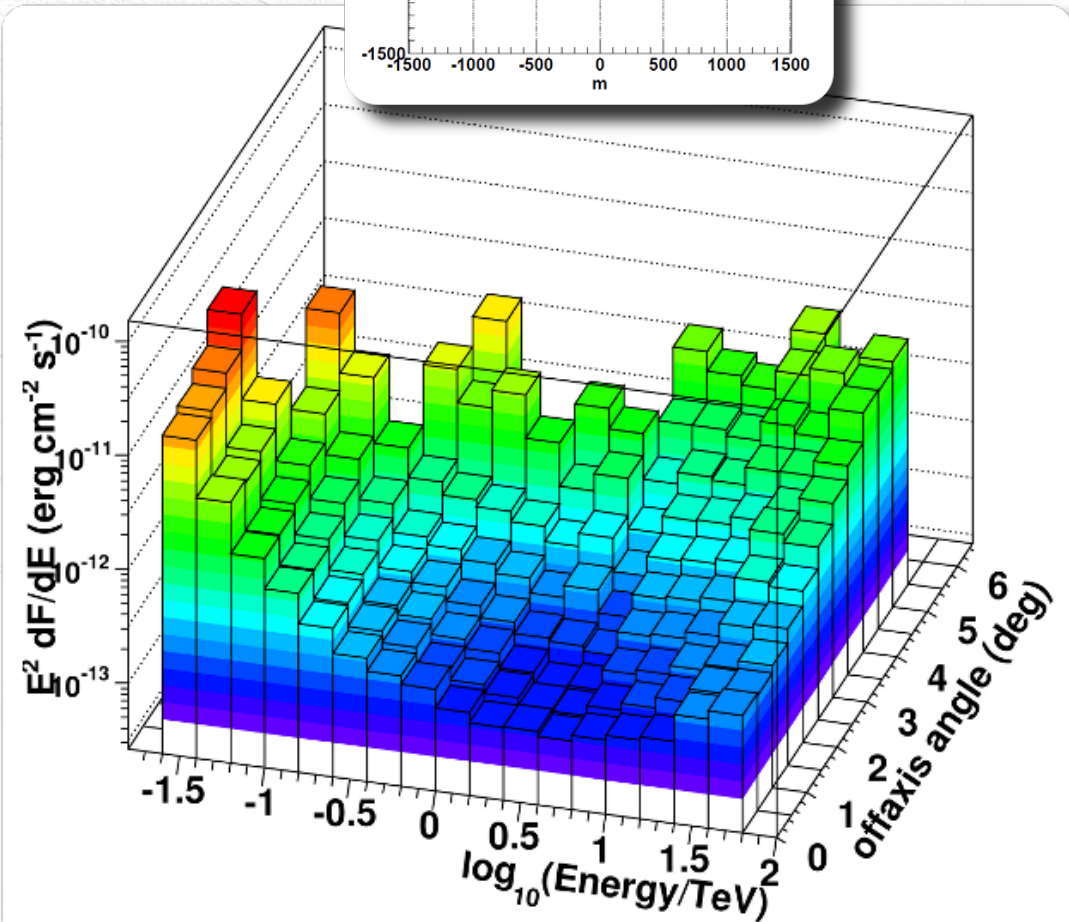
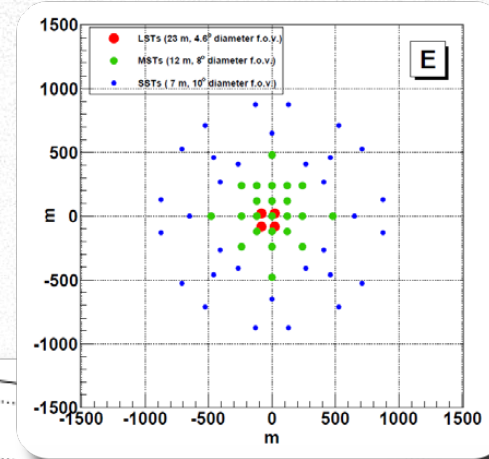
First Results

➔ analysis of simulated gamma and proton (background) showers to estimate sensitivity of possible arrays

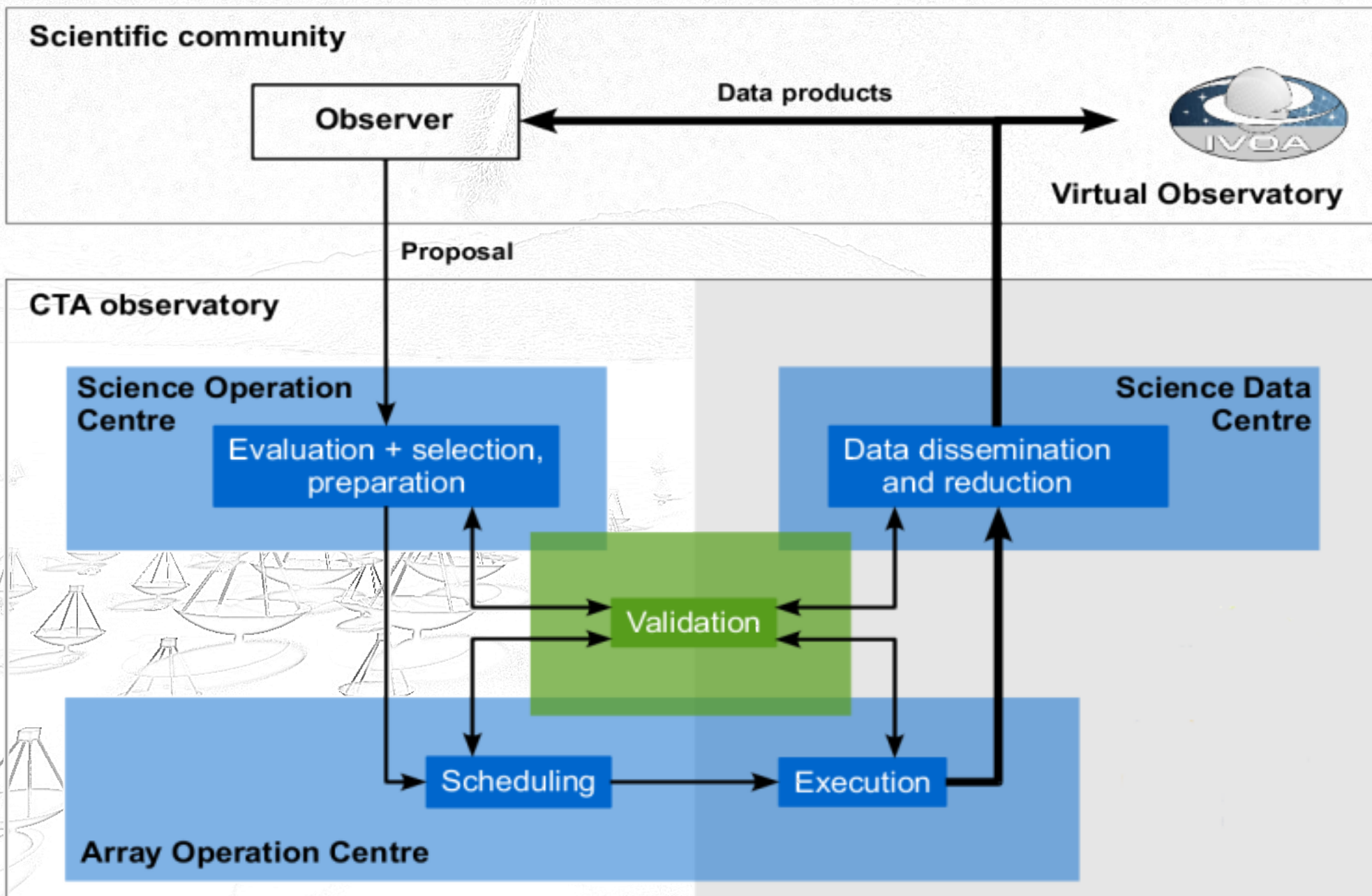
➔ sensitivity for array layout E

• [diPierro et al., ICRC 2011]

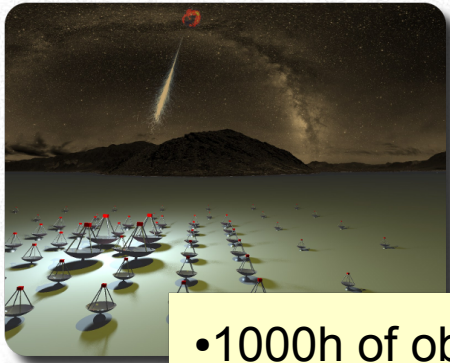
➔ input for physics working group to decide on scientific potential



Future Operational Workflow



Future CTA Data Flow



- 1000h of observation
- data stream 1-10GB/s
- some PB per year

remote site:
internet sufficient for data transfer?
build new connection?



- calibration and reconstruction
- 1h raw data needs ~700CPU*days
- file size: ~10% of raw data

first analysis

Virtual Observatory

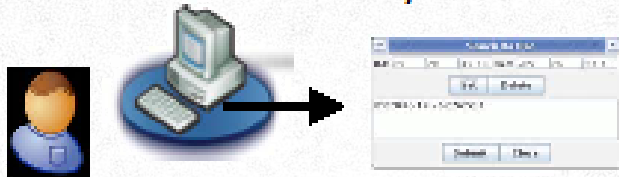


available
to public

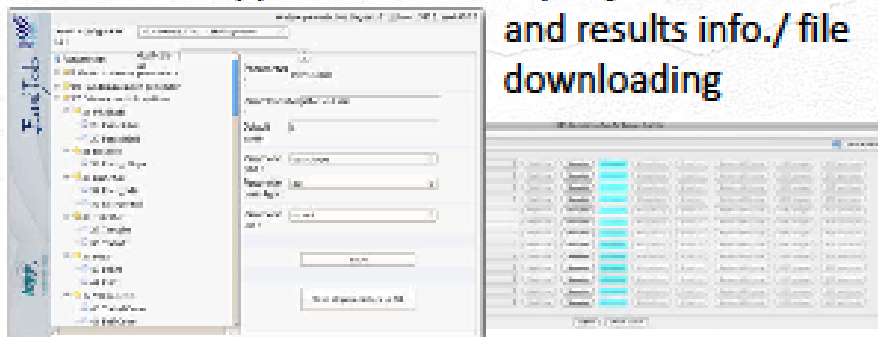


Science Analysis System

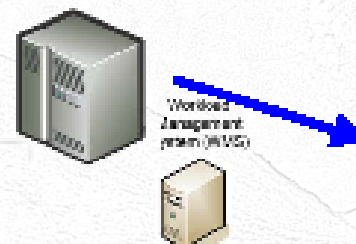
User Web Client and/or VM for data searching



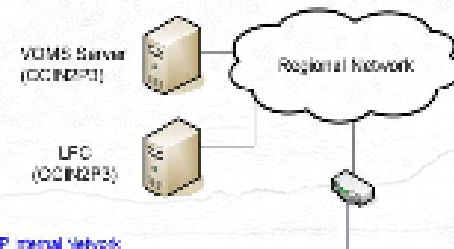
Client application for analysis job submission and results info./ file downloading



Web Server and Grid jobs broker

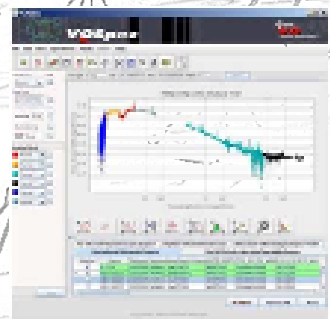


Grid jobs management



High-level data request

FITS and VObs data access



FITS and VO compliant results
VObs tools (VOSpec, Aladin)

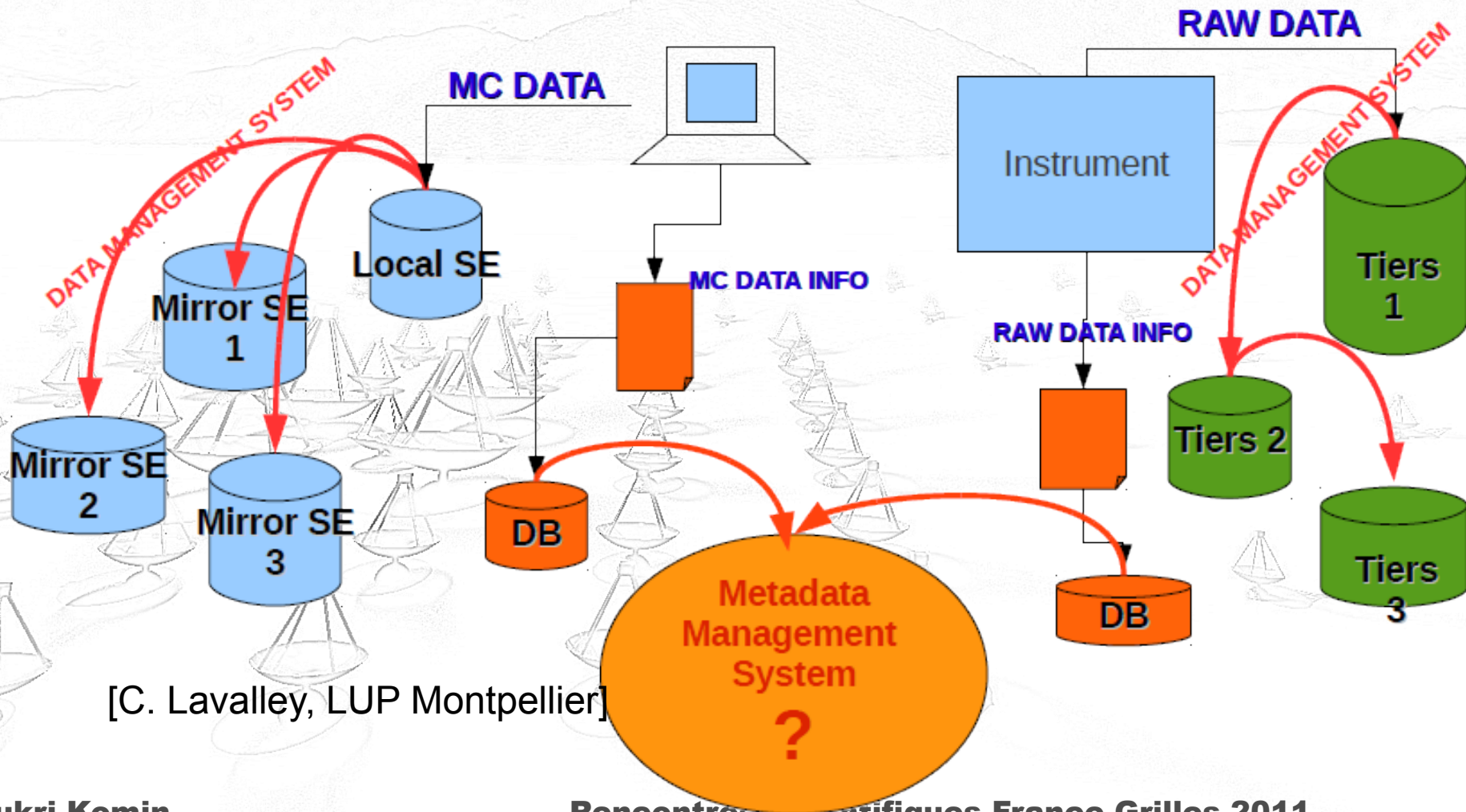
Interface layer with archive DB

Science Archive, DB and MetaDB systems, notification and results uploading



Meta Data Management

- ⇒ different data bases in use, dependencies between files
- ⇒ need Meta Data Management
- ⇒ Atlas Meta Data Interface (AMI) under evaluation



[C. Lavalley, LUP Montpellier]

Summary

⇒ Cherenkov Telescope Array

- future large scale observatory for gamma-ray astronomy
- preparatory phase

⇒ CTA Computing Grid

- currently massive Monte Carlo production

⇒ future plans: set up full analysis chain

- Science Gateway: user interface for data access and analysis
- meta data management
- data transport (in particular the raw data)
- computing: Grid, cloud or something else?

⇒ CTA grid computing at the beginning, we're looking for solutions.

⇒ Many thanks to the sites supporting CTA.