

CTA

Cherenkov Telescope Array Computing Grid and e-infrastructures

Giovanni Lamanna

LAPP - Laboratoire d'Annecy-le-Vieux de Physique des Particules, Université de Savoie, CNRS/IN2P3, Annecy-le-vieux, France



- Introduction: What is CTA ?
- CTACG project: CTA design study and MC production
- CTA Observatory E-Infrastructures
- CTA data access
- Perspectives
- Conclusion



IACT



VHE Gamma-ray infrastructures



Cherenkov Telescope Array

- CTA is a new project for ground based gamma-ray astronomy planned to consist of several tens of Imaging Atmospheric Cherenkov Telescopes (IACTs)
- CTA is an Astroparticle research infrastructure aimed to work as an observatory providing services making gamma-ray astronomy accessible to the entire community
- The CTA international consortium (after a Design Study phase) is currently committed in its (EC-FP7 funded) Preparatory Phase (2011-2013)



The CTA consortium

The CTA international consortium counts > 100 institutes from > 22 countries:

- 1) Armenia,
- 2) Bulgaria,
- 3) Croatia,
- 4) Czech Republic,
- 5) Finland,
- 6) France,
- 7) Germany,
- 8) Greece,
- 9) Ireland,
- 10) Italy,
- 11) Netherlands,

arXiv:1008.3703v1 [astro-ph.IM] 22 Aug 2010

- 12) Poland,
- 13) Spain,
- 14) Switzerland,
- 15) UK,
- 16) Denmark,
- 17) Sweden,
- 18) Argentina,
- 19) Japan,
- 20) Namibia,
- 21) South Africa, 22) USA.



Design Concepts for the Cherenkov Telescope Array CTA

An Advanced Facility for Ground-Based High-Energy Gamma-Ray Astronomy

The CTA Consortium

May 2010



• **CTACG (CTA Computing Grid) aims at:**

- Optimising the application of grid technologies for the CTA:
 - Simulation
 - Data processing
 - Storage
 - Offline analysis
 - Virtual Observatory interface

CTA COMPUTING GRID

Contingent motivation in 2008:

Computing resources for MC simulations in the Design Study phase



Why GRID?

- CE and SE to fulfill Monte Carlo production and data reduction requirements
- Central production and control
- Distributing data produced by "distributed" resources (Worker nodes)
- User access to computing resources and analysis tools
- A step toward a coherent CTA data management system
- (At a cost of some initial efforts) Benefit of LCG experience, technical support and mainframe (e.g. middleware, security, db, tools,...)
- Opportunity for LCG Tier2 to open up to a multi-disciplinary context for mutual exchange of experience (e.g. conjugating VObs and EGImiddleware, facing different requirements from the Astroparticle community) towards Astrop.- "Data Centers".



From CTACG to CTA E-Infrastructures



- <u>Grid infrastructure</u>: a dedicated EGEE VO named vo.cta.in2p3.fr created in 2008 and including 14 sites in 4 countries (France, Germany, Spain, Poland)
- Resources overview

Country	Site	Available Storage (TB)	Available CPUs
	IN2P3-LAPP	15	760
Eropoo	IN2P3-CC	280	4745
France	GRIF-IRFU	10	4670
	MSFG-OPEN	10	104
Cormony	MPPMU	10	856
Germany	DESY-ZN	10	672
Spain	pic	7	1365
Spain	CIEMAT-LCG2	1	824
Deland	CYFRONET-LCG2	100	1984
Poland	CAMK	5	0
Total	All CTA sites	448	15980

<u>Available software</u>: Corsika + sim_telarray

- \Rightarrow ~ 450 TB available (280 TB used)
- ⇒ fulfills MC requirements:
 > 400 TB SE, > 300 CPUs CE



• Requirements :

- Automated/permanent tool
- Generic and self consistent tool
- Embedded monitoring, logging and bookkeeping systems
- Tool based on already tried and tested EGEE Grid tools
- Easily configurable tool to run simulations, analysis or other tasks
- Smart management of Grid resources, according to available resources
- Monitoring of sites, tasks, jobs, output files
- From a user point of view:
 - No specific Grid knowledge needed to run the jobs and get information and output files from the jobs



CTACG Architecture





Based on GANGA

(Gaudi/Athena aNd Grid Alliance: http://ganga.web.cern.ch)

 Grid front-end for job definition and management in Python: developed, used and maintained by the LCG (LHC Computing Grid) community

- CTA Grid interface developed at LAPP (by S.Elles and O.Arnaez - Atlas LAPP group)



• Main features:

- Runs as a daemon for CTA
 - ⇒ infinite loop that automatically launches all tasks defined in the central database
- Automated start of new tasks once defined
- Stop, restart, delete jobs/tasks possible at any moment on demand
- Embedded Grid resources management tool
- Logging, bookkeeping, monitoring systems
- Current status: applied to MC production:
 - Centrally done at LAPP
 - Will then be open to all users



CTACG - User interface



Task configuration



- Web interface developed in Java at LAPP enabling for each software:
 - Software parameters creation (lists, one by one)
 - Setting of values for these parameters (some of them mandatory)
 - Cloning of a software configuration (ex: need to add a few parameters when new software released)
 - Cloning of all parameters values (ex: need to run the same software with different input values)
- Central database automatically updated upon changes
 2 available levels:
 - Administrator level with full access to the tool
 - User level allowing to change only a few parameters (chosen by the administrator)



Create or clone a software configuration

🕲 CTA Production Software Manager - Mozilla	Firefox	
Eichier Édition <u>A</u> ffichage <u>H</u> istorique <u>M</u> arque-page	s <u>O</u> utils <u>?</u>	
CTA Production Software Manager 🛛 🕂		-
		🎴
CTA PRO	DUCTION : SOFTWARE CONFIGURATION	
		Context
Software Parameter	Properties	Rarameter Definition
🕈 📑 Softwar	Properties	Parameter Instanciation
Create Software Config.		✓ Expert Mode
Clone Software Config.		
🗐 🔤 🔤		
Modification Group Composition		
Delete selected Parameter Group		
Create Parameter Item		
Create new Parameter		
Delete selected Parameter		
Create Unit		
Delete selected Unit		
Create Type Definition		
Delete selected Type Definition	State : LOCKED	
Cherenkov emission parameters	Prod. : p7(linked to config. : CORSIKA)	
Cebugging and output options	Created by p1 at : 2010-04-01 09:59:25.0	
💡 🛋 Parameters Configuration	State : LOCKED	
— 🤷 ARRANG[float]	Prod. : p1(linked to config. : <u>CORSIKA</u>)	
	at : 2010-04-01 14:36:23.0	
	State : LOCKED	
	Prod. : p8(linked to config. : <u>CORSIKA</u>) — Created by cecile	
bolean	at : 2010-04-01 16:32:51.0	
CERFILL boolean J	Prod. : TestProd/ linked to config. : CORSIKA)	
— 🍫 CERSIZ[float]	Created by PP	
- 💫 COMOUT[boolean]	State : IN CONSTRUCTION	Save
	Production Instances	Cancel
2010-4-2 11:2:41 DB Server contacted : lappweb02.in2p3.fr 2010-4-2 11:2:41		
2010-4-2 11:10:26 Select the instance to load. (Right button on :	elected software) or create a new one	
2010-4-2 11:10:43 current instance unioaded		
Tarriné		
I CI IIIII I C	U:6	



Create parameters and groups of parameters in the selected software configuration

CTA Production Software Manager	*	
	CTA PRODUCTION : SOFTWARE CONFIGURATION	
Software Parameter	Properties Properties Descent black i f	-Context Parameter Definition
🕈 💼 Parameter Groups Configur	Destruction :	Expert Mode
- 🦄 RUNNR EVTNR NSI <u>=</u> ∞ 戦 seed	Add a new constraints	
 Primary Particle Option Site Specific options Core Range 	Load Description	
Y 🐔 Telescope positions, for	Mandatory K Expert View only Validate Cancel	
କ 🏀 Interaction flags କ 🍢 Cherenkov emission pa		
Parameters Configuration	State : LOCKED Prod. : p7(linked to config. : CORSIKA) Created by p1	
	at : 2010-04-01 09:59:25.0 State : LOCKED Prod. : p1(linked to config. : <u>CORSIKA</u>) Constitution of the p1	-
boolean	Created by pr at: 2010-04-01 14:36:23.0 State: LOCKED Prod.: p8(linked to config.: CORSIKA)	
- To CERFIL[boolean] - To CERSIZ[float]	Created by cecile = = = = = = = = = = = = = = = = = = =	Ξ
- 🕠 COMOUT[boolean]	Prod. : TestProd(linked to config. : <u>CORSIKA</u>) Created by PP at : 2010-04-01 18:05:48.0 State : IN CONSTRUCTION	Save
	Production Instances	Cancel
110-4-2 11:2:41 DB Server contacted : lappweb02.ir 110-4-2 11:2:41 110-4-2 11:10:26 Select the instance to load. (Right l 110-4-2 11:10:43 current instance unloaded 110-4-2 11:13:2 Select the instance to load. (Right h	p3.fr utton on selected software) or create a new one tton on selected software) or create a new one	
10-4-2 11:15:2 Select the instance to load. (Kight D	roll oll zelerien zollmaus) oll cleare a Liew olle	





Load, create or clone a production instance



Set values to mandatory parameters of a production instance • Enables the administrator to choose on which sites to run according to all information he has (EGEE broadcast, failed jobs on one site, ...)

er Edicion Arrichage Historique Marque-pages Qui	.clis <u>ŕ</u>	
CTACG Site Configuration Interface		
	CTACG Site Configuration Interf	ас
Site status		
IN2P3-LAPP		
⊙operational ⊂ignore ⊂down		
IN2P3-CC		
⊙operational ⊂ignore ⊂down		
GRIF-IRFU		
⊙operational ○ignore ○down		
GRIF-LAL		
○operational ⊙ignore ○down		
GRIF-APC		
🔿 operational 🔿 ignore 🔞 down		
CRIE-LIR		

• Enables the administrator to choose scripts and executables to be submitted, number and subsets of jobs to run according to available resources; to start the jobs submission.

CTACG Job Submission Interface ÷		
СТАСС	G Job Submis	sion Interface
Production parameters	Bunch size	Bunches per site
Production instance choice	0 1	IN2P3-LAPP 4 operational 2184 GB 763 CPUs
Script Directory Write here the path to the script you wish to run	○ 2	IN2P3-CC 10 operational 15728 GB 3526 CPUs
SE Data Path Write here the path where to store the output files on the SEs	03	GRIF-IRFU 2 operational 189 GB 1165 CPUs
Software Name		GRIF-LAL 2 operational 528 GB 47 CPUs
CORSIKA	0 20	GRIF-APC 2 operational 4691 GB 56 CPUs
v_03_18feb10	● 50	GRIF-LLR 2 operational 916 GB 163 CPUs
Grid Shell Filename LN PRODex10.sh	0 100	GRIF-LPNHE 2 operational 0 GB 586 CPUs
LFN Automatic Detection		MSFG-OPEN 3 operational 10625 GB 104 CPUs
⊛ γes		MPPMU 10 operational 4566 GB 0 CPUs
© no		DESY-ZN 4 operational 4566 GB 0 CPUs
		PIC 10 operational 12354 GB 72 CPUs

• Enables the administrator to select an existing production and change its status.

Eichier	Éditio <u>n</u> (Affichage	Historique	<u>M</u> arc	que-paç	jes <u>C</u>	utils	2		
🗋 СТА	ACG Site C	onfigurat	ion Interfa	ice	+					
									CTACG Production Interface	
Prc	70 p6 (sto 24 p5 (do 59 p5 (do 69 p6 (rea 70 p6 (sto 71 p6 (sto 72 p7 (sto	n name pped) ne) ne) ady) pped) pped) pped)	e and cu	irrent	t stat	us—			Production new status required Cmd_none Cmd_stop_submission_and_kill_jobs Cmd_stop_submission Cmd_restart_failed Cmd_restart_from_zero	
									Valid	

• Enables the administrator to select a subset of jobs among an existing production and change its status.

🕹 CTACG Bunches Interface - Mozilla Firefox		
Eichier Édition Affichage Historique Marque-pages Outils ?		
CTACG Bunches Interface		
	CTACG Bunches Interface	<u>^</u>
Production #70 (cmd_none)		
Bunches number and current status	Bunch new status required	Bunch new site required
#662 (ready) IN2P3-LAPP #662 (ready) IN2P3-LAPP #663 (ready) IN2P3-LAPP #664 (ready) IN2P3-CC #665 (ready) IN2P3-CC #666 (ready) GPIF-APC #667 (ready) GPIF-APC #668 (ready) MSFG-OPEN	 cmd_none cmd_stop_submission_and_kill_jobs cmd_stop_submission cmd_restart_failed cmd_restart_from_zero 	 IN2P3-LAPP operational 2184 GB 763 CPUs IN2P3-CC operational 15728 GB 3526 CPUs GRIF-IRFU operational 189 GB 1165 CPUs GRIF-LAL operational 528 GB 47 CPUs GRIF-APC operational 4691 GB 56 CPUs GRIF-LLR operational 916 GB 163 CPUs GRIF-LPNHE operational 0 GB 586 CPUs MSEC-ODEN operational 10625 CP 104 CDUs
Terminé		0:3 💥 💷 🔂 😥



Monitoring



- Display of site status, storage and CPUs available at each site
- Detailed job status per site



- Display of tasks status (global and per site)
- Detailed job status per site

CTA-CG - Task status	4						
-							
				CTA-CG platform - Tasks statu	s - LAPP		
							Wed Apr 7
p7		ld:72	ready	CORSIKA v_03_18feb10			Wed Apr 7 15:57:04 2010
global job list	1			56	6 63	89%	Detailed job status
IN2P3-LAPP	6				6	10%	scheduled waiting funning done faile
IN2P3-CC	18				18	29%	
GRIF-LAL	3				3	5%	
MSFG-OPEN	12				12	19%	
MPPMU	3				3	5%	
DESY-ZN	1 8				9	13%	
PIC	6				6	0%	
CYFRONET-LC62	6				6	10%	
p6		ld:71	stopped	CORSIKA v_03_18feb10			Thu Apr 1 19:35:66 2010
p6		ld:70	stopped	CORSIKA v_03_18feb10			Wed Apr 7 11:50:27 2010
global job list				30	30	0%	Detailed job status
IN2P3-LAPP	10				10	0%	scheduled waiting running done faile
IN2P3-CC	10				10	0%	
GRIF-APC	5				5	0%	
MSFG-OPEN	6				5	0%	
DESY-ZN	5				5	0%	
p6		ld:69	stopped	CORSIKA v_03_18feb10			Tue Apr 6 16:28:18 2010
global job list				75	75	0%	Detailed job status
IN2P3-LAPP	6				6	0%	scheduled waiting running done faile
IN2P3-CC	15				15	0%	
GRIF-IRFU	3				3	0%	
GRIF-LAL	6				6	0%	
GRIF-APC	3				3	0%	
ORIF-LER	3				3	0%	
GRIF-LPNHE	3				3	0%	
MSFG-UPEN	3				3	0%	
DESYZN	6				8	0%	
PIC	0				0	0%	
CIEMAT-LCG2	3				3	0%	
0121071 2002						~ ~	

- Display of jobs status within a task
- Detailed list and status of output files in SEs



Access to result files



Files download

🕙 CORSIKA dashboard	1 - CORSIKA showers browser - Mozilla Fi	refox		
CORSIKA dashboard	d - CORSIKA show 🔸			-
	[<u>Home</u> <u>Shower download</u>	Dashboard Recent jobs Resubmit Sites	<u>s Scripts Shifter zone About</u>]	^
	(Get list of CORSIKA sho	wers	Additional info N.B.: Select properties of
	Select SHOWER parameters			wish to study. A
	LIBRARY	ENERGY	ZENITH ANGLE	list of LFN will be
<u>Cherenkov</u> <u>Telescope</u> <u>Arrav</u> <u>Home</u>	Proton::library_gr0 🛆	en0 en1 en2 en3 en4 en5 en6 en7	th0	downloaded. N.B.: Note, that the LFN list will be saved in a file of your choice. The full LFN set filesize can be of the order of 10
<u>Get showers</u> <u>Miscellaneous</u>	Select file to list	en17.500: Energy of primary particle is log10(E/eV)=17.5 Small files: .info, .input, .lst, .tab, .md5.sum Logs: .job.out, .job.err	th60.000: Zenith angle of primary particle is theta = 60 degrees Getthe LFN list!	MiB!
				×
Terminé			0:6 🔀	

- Search engine through a web interface:
 - Users can list files matching the research criteria
 - In the future, they will be able to download these files
- **Status**: in progress (inspired by Auger experience)

CTACG dashboard

- Inspired by the Auger dashboard (4th EGEE User Forum Catania)
- Used to access and run all the previous tools from a unique web page

Status: operational, not yet online



The CTA Observatory main logical units :

- Science Operation Centre: organisation of observations
- Array Operation Centre: the on-site service
- Science Data Centre:
 - Software development
 - Data analysis
 - Data reduction
 - Data archiving
 - Data dissemination to observers



Total expected data volume from CTA: 1 to 10 (?) PB per year (main data stream for permanent storage is of the order of 1 (10 ?) GB/s) MC requirements: tens of CPU years, hundreds of TB

Existing ICT-based infrastructures, such as EGEE and GEANT, are potential solutions to provide the CTA observatory with best use of e-infrastructures.



Development of an integrated application of e-infrastructures like EGEE-Grid, GÉANT and IVOA for the architecture of the CTA observatory and...

Deployment of end-to-end services in support of a wide scientific community for open data access for CTA and VHE gamma ray astronomy.

EGEE-Grid infrastructures and EGEE-Grid middleware for distributed data storage, data analysis and data access are considered the most efficient solution for the CTA e-infrastructure (CTACG).





A dedicated CTA Work Package

1) Data Management System and Data Pipelines

(deliverables: framework, data applications and data products) The data management system begins at the data acquisition interface and flows through to the data products accessed by end users

2) Data Management Facilities

(deliverables: network, computing model, global system and specific roles of facilities) The data flow, on the way, moves through different facilities supporting data management according to different tasks, as well as end user sites that may conduct science using CTA data.

3) Data Access

(deliverables: services for data and computing sharing and public openness) End users and end user sites that may conduct science using CTA data require different level of data access and

different tools, resources and services to be part of the community (VHE gamma ray gateway). Further more users want to pipeline resources on their own computing infrastructure.

-> Development for USER REQUIREMENTS



A possible workflow





... toward a "VHE γ-ray Science GATEWAY"

VSG, through a Web-interface (... Virtual Machine) will provide:

- A complete, portable and easy to configure user environment for developing and running CTA data analysis locally and on the Grid (login, proxy on demand, selecting data level, data sets, software, and computing);

- Grid services for jobs and resources monitoring, archive access, telescope and observations monitoring, and support to scientist to port their software to the GRID;
- MC datasets and open production for deeper analysis of intermediate level data, On-the-fly reprocessing from the Archives Instruments calibration monitoring and trend analysis;
- Astronomical tools and services integrating VObs standards and gLite-EGEE middleware for analysis of high level data sets;
- Tools and IVOA interfaces for multi-wavelength A&A analysis;
- Forum and scientific documentation for an up-to-date scientific outreach and strengthening of the VHE gamma ray community.





This project conjugates the long-term presence and experience in EGEE/EGI of the particle physics community and the activities of the astronomical community mainly devoted to the Grid solutions for public access and archive of observatory data.

Major challenges in CTA e-infra. applications shared between two communities :

A&A

- Astrophysical data unified content descriptors
- Vocabulary for astronomical quantities extended to VHE gamma rays
- Astronomical catalogs and metadata
- -Standard grammar for database queries

PP

- Hierarchical data model definition and corresponding implementation of software and computing solutions (for processing and storage)
- Evolution towards open software and data access
- Development of Grid Analysis platform at all levels (Virtualization)
- Implementation of EGEE Grid services for the community.



From low level data To high level data and (MWL) data archive



Annex (not related to CTA)

CLOUD

... but which one ?





HoMe project

Partners : LAPP, LPSC, SBT/INAC & SYMME

Computing work package : GRACIA (Grid And Cloud Infrastructure for the Alps)

evolve from current EGI grid and high performance computing infrastructure to cloud flexible services to better meet the research community users and socio-economic partners computing requirements

Computing is seen as services :

- Software as a service (SaaS): complete application offered as a service on demand.
- Infrastructure as a service (IaaS): delivers basic storage and computing capabilities as standardized services over the network (based on a on-demand virtual images deployment system)
- Platform as a service (PaaS): encapsulates a layer of software and provides it as a service that can be used to build higher level services





- CTACG:
 - Current developments fulfil requirements
 - CTA MC Production successfully tested and running
- Perspectives:
 - Further adaptation of these tools to other CTA software (e.g. data analysis)
 - Extension of Ganga tools
 - Provide public access to data and software
- "CORE": CTA Observatory E-infrastructures:
 - Development of the CTA Data Management System
 - Data access from low (including computing resources access) to high level (interface with IVOA and running MWL VObs data on EGEE-middleware)
 - Towards a distinction between Data Operation Centre and Science Data Centre (more virtual than real, e.g. interactive services to be used by any "classical" data centre and ... by a CTA SOC ?)

• Further questions/comments:

- http://lappwiki01.in2p3.fr/CTA-FR/doku.php?id=cta-computing-grid-public
- ctacg-support (at) lapp.in2p3.fr

