

« CTACG – CTA Computing Grid »

Cecile Barbier (1), Sabine Elles (2), Nukri Komin (3), Giovanni Lamanna (4), Luisa Arrabito (5), Claudia Lavalley (6), George Vasileiadis (7), Bruno Khelifi (8), Jean-Philippe Lenain (9)

- (1) *cecile.barbier@lapp.in2p3.fr, LAPP – CNRS/IN2P3*
(2) *sabine.elles@lapp.in2p3.fr, LAPP – CNRS/IN2P3*
(3) *nukri.komin@lapp.in2p3.fr, LAPP – CNRS/IN2P3*
(4) *giovanni.lamanna@lapp.in2p3.fr, LAPP – CNRS/IN2P3*
(5) *luisa.arrabito@lupm.in2p3.fr, LUPM – CNRS/IN2P3*
(6) *claudia.lavalley@lupm.in2p3.fr, LUPM – CNRS/IN2P3*
(7) *george.vasileiadis@lupm.in2p3.fr, LUPM – CNRS/IN2P3*
(8) *bruno.khelifi@llr.in2p3.fr, LLR – CNRS/IN2P3*
(9) *jean-philippe.lenain@obspm.fr, LUTH – CNRS/INSU (now at LSW-Heidelberg)*

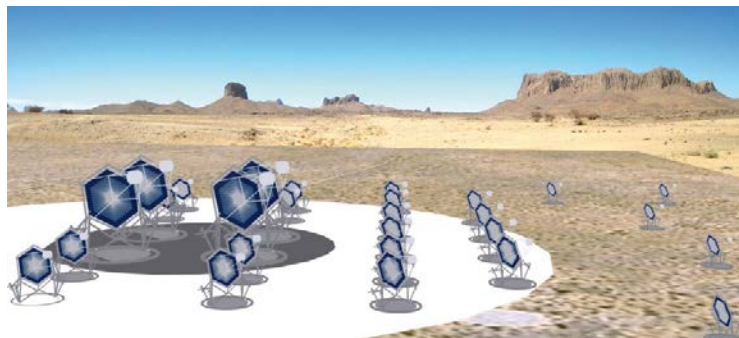


Figure 1: Vue artistique de CTA (CTA artistic view).

Overview:

In the field of very high energy gamma-ray astronomy, Europe and France in particular hold a clear leading position with the current Cherenkov instruments (e.g. H.E.S.S.). The Cherenkov Telescope Array (CTA) [1] – an array of many tens of Cherenkov telescopes deployed on an unprecedented scale – will allow the European scientific community to remain at the forefront of research.

One of the challenges to design the CTA observatory is to handle the large amounts of data generated by the instrument and to provide simple and efficient user access at any level and according to astrophysical standards in order to serve the data and the software for data analysis to the physics community. The high data rate of CTA together with the large computing power requirements for Monte Carlo simulations, fundamental tool for data selection and calibration, demand dedicated computer resources which can be well handled through the Grid approach.

During the Preparatory Phase of the CTA project a massive production of CTA Monte Carlo simulations for the investigation of physics of atmospheric cosmic-ray showers (through the CORSIKA package) as well as for the investigation of the Cherenkov telescope response and performance as a function of detectors and lay-out configurations, is taking place since 2008 within a dedicated EGI CTA Virtual Organization (VO). Simulations have been provided to the full CTA consortium making available the extensive CTA-VO computing resources (thousands of CPU equivalents and more than 500 TB data storage).

EGI-Grid infrastructures and EGI-Grid middleware for distributed data storage and data access are considered one of the most efficient potential solutions for the CTA e-infrastructure. A feasibility study of applications of Grid solutions for CTA is in progress within a dedicated CTA Computing Grid (CTACG) [4][5][6] project. CTACG is contributing in three main activities: 1) Data management system and data pipelines; 2) Data management facilities for any major logical unit of the observatory. 3) Data access: developing services for data and computing sharing and public openness.

A series of applications and services are under development and some already deployed, among these: - Integrated applications “EasIJob” based on GANGA (Gaudi/Athena aNd Grid Alliance) for workflow configuration, grid jobs submission (figure 3.1), monitoring control tools and services and bookkeeping system; - User interface for data browser; - A meta-database system for instrument response function simulation configurations management; - DIRAC interface for end-user analysis interface (figure 3.2) and data catalogue interface, possibly compliant with high level astronomical data format (e.g. FITS and Virtual Observatory). The astroparticle community committed in CTA demonstrates to merge the experiences from

the particle physics (e.g. EGI and WLCG), astrophysics (e.g. Virtual Observatory), and major European e-infrastructure (e.g. GEANT, EGI and Euro-VO) and to explore frontier Information and Computing Technology (ICT) developments.

Enjeux scientifiques, besoin de la grille :

Le projet CTA (Cherenkov Telescope Array) [1] (figure 1) a pour objectif la construction et l'exploitation d'un réseau de plusieurs dizaines de télescopes Cherenkov déployés à une échelle sans précédent. Cet instrument permettra à la communauté scientifique européenne de rester à la pointe de la recherche dans le domaine de l'astronomie gamma de très haute énergie. CTA ouvrira une nouvelle ère dans l'astronomie gamma de très haute énergie, grâce à des performances très supérieures à celles des projets actuellement en cours dans ce domaine : un gain d'un facteur 10 en sensibilité, ainsi qu'une grande couverture spectrale, une meilleure résolution angulaire et temporelle et une couverture complète du ciel.

Le projet CTA est l'un des projets phares de la décennie qui s'ouvre en astronomie et astrophysique. On peut se référer pour s'en convaincre aux feuilles de route établies en 2008 par le European Strategy Forum on Research Infrastructures (ESFRI) [2] et l'ASTroParticle ERANet (ASPERA) [3].

Enfin avec CTA, l'astronomie gamma de très hautes énergies évoluera vers un statut d'observatoire où la communauté astronomique internationale pourra soumettre des demandes d'observation et accéder aussi bien aux données qu'aux logiciels d'analyse [5][6].

L'un des enjeux majeurs de l'observatoire CTA sera donc la gestion de la grande quantité de données générées par le réseau de télescopes et l'accès aux données de tous niveaux qui doit être simple, efficace et en accord avec les standards utilisés en astrophysique. Le principal flux de données devant être stockées de manière permanente est estimé à des valeurs entre 1-10 GB/s, ce qui correspond à un volume de données allant de 1 et 10 PB par an.

Pendant la phase préparatoire du projet CTA et dans le but d'optimiser les solutions techniques pour obtenir la meilleure configuration possible pour l'observatoire CTA, les performances du réseau de télescopes ont demandé des études détaillées de simulations de Monte Carlo (MC). Ces simulations demandent une grande quantité de temps de calcul (supérieur à 10 CPU/ans pour chaque configuration possible du réseau) pour: plus de 10^{10} gerbes atmosphériques induites par des protons cosmiques simulés; le traitement de toutes les interactions des particules secondaires avec l'atmosphère et, pour chaque gerbe; le suivi des trajectoires des particules et de la lumière Cherenkov associée; calcul de la projection de cette lumière dans le système optiques des télescopes; la simulation de la réponse des photo-détecteurs et de l'électronique associée.

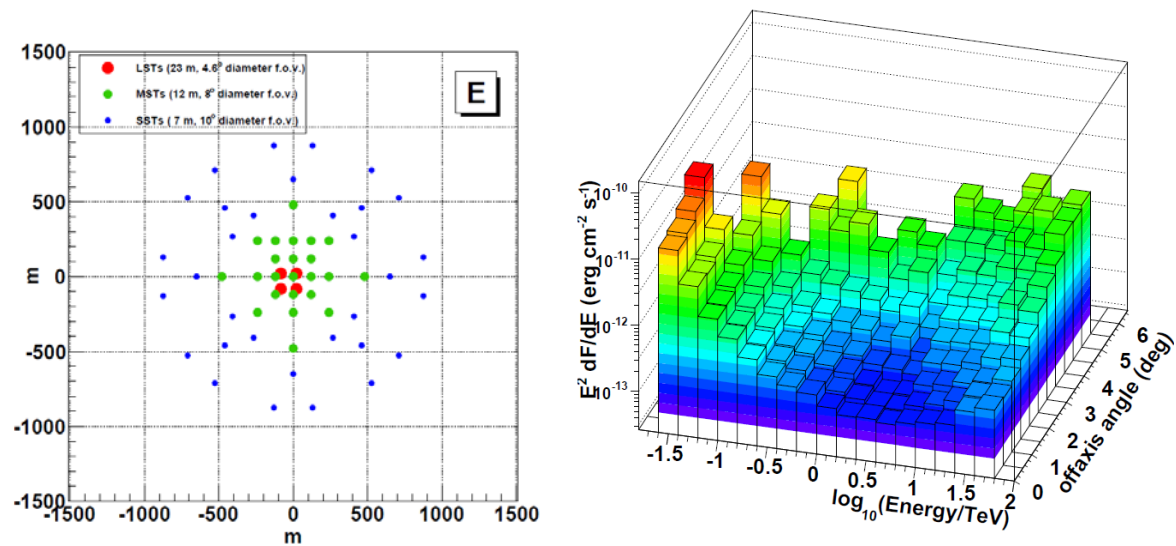


Figure 2: a) Une des configurations possibles (E) du réseau de télescopes CTA à l'étude (*Layout of one of the CTA configurations (E) under study*). b) Sensibilité simulée et analysée sur la Grille de calcul pour le réseau E. L'histogramme montre la dépendance du flux minimale (en $\text{erg cm}^{-2}\text{s}^{-1}$ et pour une détection significative à 5 sigma pour chaque intervalle) de l'énergie et de l'angle de décalage par rapport au centre des caméras (*Simulated CTA gamma-ray sensitivity obtained from MC analysis on the Grid. The plot shows for array E of CTA the minimal flux (in $\text{erg cm}^{-2}\text{s}^{-1}$) for a 5 sigma detection in each bin with respect to the energy and offset angle to the centre of the cameras.*)

Etant donné le débit de données élevé et l'importance de la puissance de calcul requise pour les simulations MC, outil fondamental pour la sélection de données et pour la calibration des instruments Cherenkov, la solution grille EGI (en termes de ressources, infrastructures et middleware) s'est révélée la plus adaptée. Les besoins en simulations massives de MC de CTA ont demandé plusieurs milliers de CPU équivalents et plus de 500 TB d'espace disque pour le stockage. Cela a permis au consortium international CTA d'avancer dans les phases de design et préparation de l'observatoire [5][7]. L'exemple d'un résultat préliminaire des études de MC est montré dans la figure 2.

Développements, déploiement sur la grille :

Une étude de faisabilité pour l'utilisation des technologies grille dans le cadre des problématiques propres à CTA est en cours au sein d'un projet dédié : le projet CTA Computing Grid (CTACG) [4]. La mission de ce projet est de concevoir et de maintenir une infrastructure pour le stockage et le traitement de données pour le consortium CTA au cours de différentes phases préparatoires [5][6] et plus tard pour la communauté scientifique internationale. Le projet CTACG a déjà prouvé que l'application des technologies grille dans le cadre des productions massives des simulations MC est tout à fait adaptée.

Une Organisation Virtuelle CTA (VO CTA) a été créée en 2008 sous l'initiative du groupe français CTA du LAPP et en collaboration avec le Centre de Calcul de l'IN2P3 (CCIN2P3). A ce jour la VO CTA est supportée par 14 sites répartis en 5 pays (France, Allemagne, Espagne, Pologne et Grèce) et elle compte environ 60 membres appartenant à plus de 20 instituts provenant de plus de 10 pays différents.

Le projet CTACG a bénéficié des infrastructures et des applications grille déjà déployées au LAPP dans le cadre de WLCG (le LAPP est centre Tier-2 WLCG), ce qui a facilité l'organisation, la gestion et le support de la VO CTA et l'établissement d'une synergie majeure avec le CCIN2P3, lequel fournit un support critique en termes de services et d'infrastructures [8] [9] [10].

Outils, résultats et difficultés rencontrées :

Les ressources disponibles dans les sites d'EGI supportant la VO CTA ont permis de simuler plus de 10^{10} gerbes atmosphériques et diverses configurations de réseau de télescopes. Les analyses de ces simulations sont en cours et des résultats préliminaires font déjà l'objet de publications [5][7]. La mise en production de l'infrastructure de simulation pour CTA dans la grille de calcul a demandé des efforts principalement dans : l'adaptation du logiciel à la grille ; l'organisation des catalogues pour l'archivage des résultats ; l'interface avec une base de données de définition des paramètres de configuration du logiciel de simulation. Les seules difficultés rencontrées concernent 15% des « jobs » de production MC, dues à une nécessité de mémoire vive (RAM) supérieure à 4 GB pour les nœuds de calcul des sites d'EGI. Cette limite a été surmontée grâce à quelques ressources spécialement dédiées et mises à disposition par de plus grands centres de calcul (e.g. CCIN2P3 et CYFRONET) couvrant ce besoin en mémoire et donc capables de répondre aux spécifications.

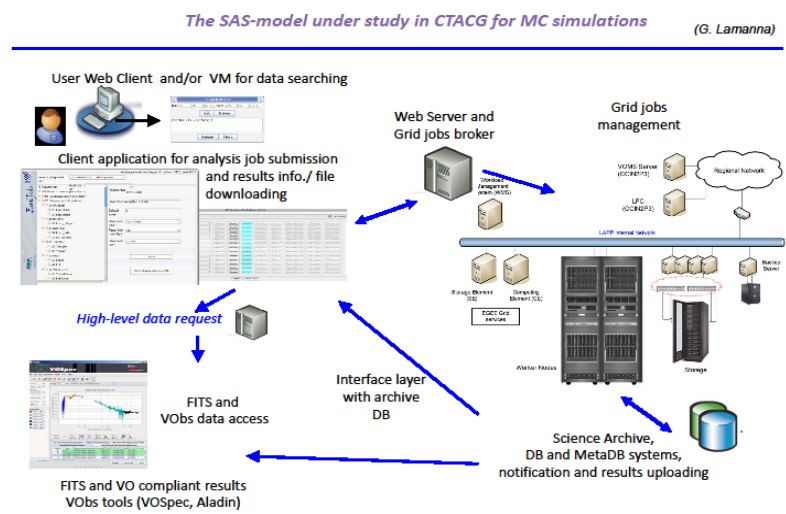


Figure 3: Vue opérationnelle du SAS basée sur l'exploitation de ressources fédérées de type Grille à l'étude dans CTACG et en application pour les simulations de type Monte Carlo (*The operational view on a Grid-CTA-Scientific Analysis System for MC simulations under study and development within CTACG*).

Le projet CTACG a aujourd'hui pour objectif de développer une série d'applications de façon à étudier l'approche grille appliquée à un Système d'Analyse Scientifique – SAS (figure 3) dédié aux simulations MC et éventuellement à l'analyse future des données.

Un ensemble d'applications et de services est actuellement en cours de développement et pour certains cas déjà déployé [6][10], tels que : - Une application intégrée "EasiJob" basée sur GANGA (Gaudi/Athena aNd Grid Alliance) (figure 4.1), interface d'accès à la grille et aux systèmes de batch, adapté en particulier pour les productions MC de CTA, le monitoring et le « bookkeeping » ; - Une interface utilisateur de navigation et d'accès aux ressources; - Un système de structuration de métadonnées permettant la gestion de données issues des simulations MC ainsi que des différentes étapes du traitement de données de l'observatoire (développements en cours au LUPM); - Des développements dans le cadre de DIRAC (Distributed Infrastructure with Remote Agent Control) [11], utilisé comme système pour l'optimisation de l'exploitation des ressources de la Grille pour les tâches d'analyse de MC (figure 4.2); - Une interface avec le catalogue de données (compatible avec le format de données de haut niveau utilisé en astronomie, i.e. FITS et Observatoire Virtuel).

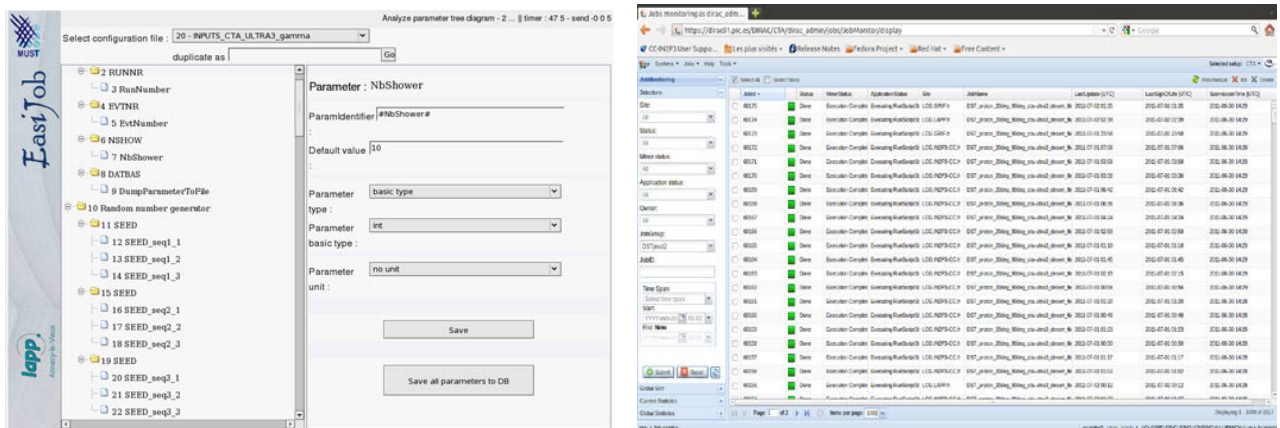


Figure 4.1: Interface web d'une des applications EasiJob développée pour CTA: sélection de la configuration du logiciel des simulations Monte Carlo et son interface de soumission (*Web interface of one of the EasiJob applications developed for CTA: software and workflow configuration interface selection for Monte Carlo simulations submitted to the Grid*). **Figure 4.2:** Service web basé sur le système DIRAC et développé pour CTA pour l'accès et l'interface aux jobs d'analyse et leurs résultats (*Web service based on DIRAC and developed for CTA devoted to user data access and interface to data analysis jobs on the Grid*).

L'atout de la communauté de physique des astroparticules dans CTA réside dans sa faculté à associer l'expérience des physiciens des particules (EGI et WLCG) et des astrophysiciens (Virtual Observatory), à profiter de la maturité des majeures infrastructures européennes (GEANT, EGI et Euro-VO) et également à explorer des développements ICT de pointe. La possibilité de fédérer également au travers de l'approche grille les communautés scientifiques des pays émergents, qui sont des candidats potentiels pour l'hébergement des réseaux des télescopes de CTA, représente une valeur ajoutée aux objectifs du projet CTACG avec en outre une plus-value évidente d'un point de vue socio-économique.

Perspectives :

Le consortium CTA est désormais rentré dans une Phase Préparatoire (2010-2013), financée par l'UE par le biais d'un projet FP7. Dans cette phase, l'exploration et le déploiement des infrastructures de calcul pour l'observatoire CTA fait l'objet d'un groupe de travail dédié. Ce groupe de travail se concentre en particulier sur trois aspects principaux : 1) Le système de gestion des données et le flux des données : exploration de l'application des technologies modernes de l'information aux opérations de l'observatoire CTA (infrastructures de calcul, stockage et transfert de données). 2) Les infrastructures pour la gestion des données : étude du rôle des grands centres de calcul nationaux et des infrastructures des pays hôtes pour chaque unité logique majeure de l'observatoire : i.e. Array Operation Centre (AOC), Science Data Centre (SDC), Support Operation Centre (SOC) (figure 5). 3) Accès aux données : développement de services pour le partage et la publication des données. Les données de tous niveaux, les outils d'analyse, les

ressources et les différents services seront mis à disposition de l'ensemble de la communauté scientifique intéressée par l'analyse des données de CTA (VHE gamma ray Gateway).

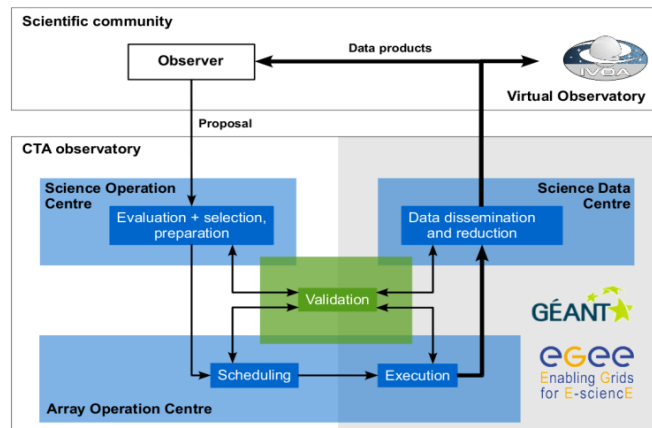


Figure 5: Schéma opérationnel de l'observatoire CTA (*Logical operational workflow of the CTA Observatory*).

References:

1. CTA: <http://www.cta-observatory.org/>
CTA @ LAPP : <http://lappwiki01.in2p3.fr/CTA-FR/doku.php?id=cta-lapp>
2. ESFRI-2008 roadmap : ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri_roadmap_update_2008.pdf
3. ASTroParticle ERANet (ASPERA) roadmap:
http://www.aspera-eu.org/images/stories/roadmap/aspera_roadmap.pdf
4. CTACG - <http://lappwiki01.in2p3.fr/CTA-FR/doku.php?id=cta-computing-grid-public>
5. "Design Concepts for the Cherenkov Telescope Array", The CTA Consortium, 2010, arXiv:1008.3703, <http://arxiv.org/abs/1008.3703v1>
6. "CTA computing Grid and e-infrastructures", G. Lamanna – Astroparticle and Computing – 1st ASPERA workshop. Lyon, October 2010.
<http://indico.in2p3.fr/contributionDisplay.py?contribId=8&confId=3845>
7. "Performance studies of the CTA Observatory" F. di Pierro et al. for the CTA Consortium, Proceedings 32nd ICRC Beijing 2011.
8. « CTA », G. Lamanna , ESFRI projects session – EGEE Conference 2009 – Barcelona
<http://indico.cern.ch/contributionDisplay.py?sessionId=12&contribId=248&confId=55893>
9. « CTACG – Cherenkov Telescope Array Computing Grid » , 5th EGEE-User-Forum Uppsala 2010 – G.Lamanna, C.Barbier, S.Elles, T. Le Flour
<http://indico.cern.ch/contributionDisplay.py?sessionId=26&contribId=157&confId=69338>
10. « CTACG – Cherenkov Telescope Array Computing Grid » EGI-User-Forum – Vilnius 2011– C. Barbier N. Komin, G.Lamanna, S.Elles
<https://www.egi.eu/indico/contributionDisplay.py?sessionId=10&contribId=75&confId=207>
11. DIRAC: <http://diracgrid.org/>