

## **Software Engineer position: data processing and workflow systems for distributed environments**

### **Missions**

The retained candidate will contribute to the development of a production system for the large-scale data processing of the Cherenkov Telescope Array (CTA) observatory. In interaction with the LUPM IT service and several members of the CTA Consortium, he/she will lead the project to bring the system developed at LUPM (CTA-DIRAC) to a production level according to the requirements of the CTA observatory.

### **Context**

CTA, Cherenkov Telescope Array [1], is a worldwide collaboration gathering about 1200 scientists from a hundred of research institutes.

The Laboratory Univers et Particules de Montpellier (LUPM [2], CNRS) is strongly involved in CTA data management and processing. In this context, LUPM has developed a prototype of a production system for the large-scale data processing of CTA, based on the DIRAC framework [3] (CTA-DIRAC instance). DIRAC is an open source software for the processing and data management in distributed environments. It is developed in a collaborative project by CERN and other research institutes (CNRS, Imperial College London, KEK, etc.).

The CTA-DIRAC instance is currently deployed in 3 computing centers in Europe and it includes a DIRAC software extension specific to CTA. LUPM team currently exploits this prototype to manage official CTA Monte Carlo simulations on the EGI grid (European Grid Infrastructure) [4]. Each simulation campaign consists of about ten thousand concurrent jobs running for several weeks and generating several hundred TB per year.

With the next-coming start of CTA observations in 2022, the goal is to bring the CTA-DIRAC instance to a production level for the processing of CTA real data. To achieve this goal, some specific developments are needed, as those concerning the components for fully automated workflows management, operation failure recovery, production validation and advanced user interfaces.

The retained candidate will integrate the LUPM IT service, composed of six engineers and he/she will contribute to all these developments as well as to the deployment of procedures for code testing and quality control. He/she will work in tight collaboration with LUPM engineers and DIRAC developers in an open source collaborative project. Moreover, he/she will work in interaction with the researchers of the CTA group of LUPM as well as with several institutes of the CTA Consortium linked to the project.

[1] <https://www.cta-observatory.org/>

[2] <https://www.lupm.in2p3.fr/>

[3] <https://github.com/DIRACGrid/>

[4] <https://www.egi.eu/>

## **Main activities**

- Manage the project of bringing the CTA-DIRAC system to a production level
  - Contribute to the DIRAC open source project for the development of advanced functionalities of workflow management
  - Design and develop advanced components of the DIRAC extension specific to CTA
- Lead the strategy for the deployment of the procedures for code testing, quality control and continuous integration
- Write technical documentation

The retained candidate will have to travel in France and abroad to present the project status at CTA Consortium and DIRAC meetings.

## **Expected background and skills**

The candidate must have a software engineer degree or a PhD degree in physics or astrophysics with consolidated experience in software development. The position is open to candidates with 1-5 years experience, but applications from young graduates or more experienced candidates will also be considered.

- Required skills
  - Python language (version 3.x): expert level
  - Linux/Unix environment: advanced level
  - DB management (MySQL, Elasticsearch): user level
  - Knowledge of tools for code management, quality control, continuous integration (gitlab, github, gitlab CI, SonarQube, Jenkins): advanced level
- Desired skills
  - Experience in distributed computing (in particular grid computing)
  - Experience in developing or using systems for large-scale data management and processing
- Complementary skills welcome
  - Containerization technologies
  - Message Queuing systems
  - Knowledges in astrophysics/astroparticle domain

Fluent English is expected (French is appreciated).

## **Location**

LUPM – UMR5299  
Université de Montpellier – Campus Triolet  
Place Eugène Bataillon – CC 72  
34095 Montpellier Cédex 05 FRANCE

## **Contract duration and salary**

2 years contract. The salary varies between 1978 € and 2517 € per month net according to the experience.

**Starting date:** From May to September 2021

**Contact:** [arrabito@in2p3.fr](mailto:arrabito@in2p3.fr)