Prototype of a production system for Cherenkov Telescope Array with DIRAC

L. Arrabito[1], J. Bregeon[1],
A. Haupt[2]
for the CTA Consortium
R. Graciani Diaz[3], F. Stagni[4],
A. Tsaregorodtsev[5]
for the DIRAC Consortium

[1] LUPM Montpellier France,
[2]DESY / Linnaeus University,
[3]University of Barcelona Spain,
[4]CERN ,
[5]CPPM Marseille France

Summary

The Cherenkov Telescope Array (CTA) [1] is the next generation instrument in the field of very high energy gamma-ray astronomy. The expected data volume is of the order of several PB/year and the corresponding processing needs are also very large. In order to manage the off-line data processing in a distributed environment, CTA has evaluated the DIRAC system, as base framework for the CTA production system. After two years of successful exploitation of the CTA-DIRAC prototype, the next step consists in developing a fully automatized execution of the CTA workflows. For this purpose we are currently evaluating and further developing the so-called DIRAC Transformation System, which offers very interesting functionalities to achieve this automatisation.

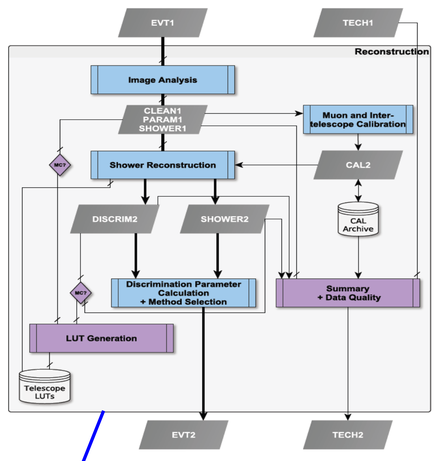
CTA Pipelines and the DIRAC Transformation System

Motivations:

- CTA uses several complex pipelines built within a 'pipeline framework'
- PB/year scale of data to be processed
- Need to automatize pipeline execution

→Use of DIRAC Transformation System to handle 'productions' (collection of identical tasks with a varying parameter)

Reconstruction Pipeline



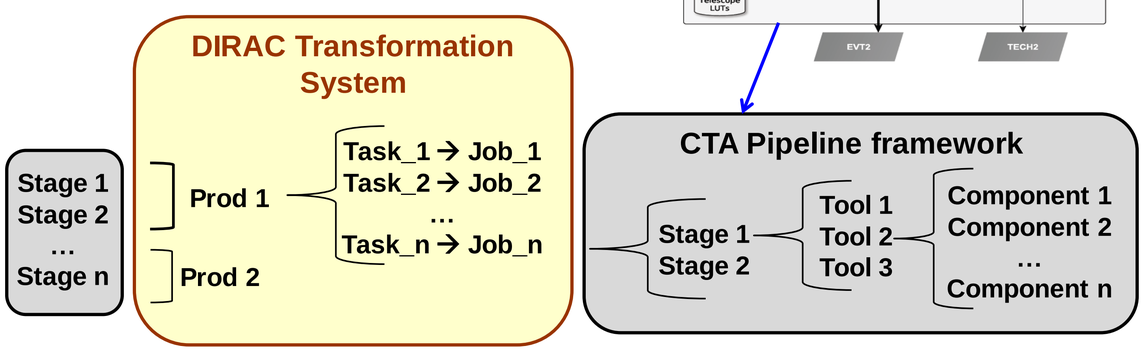


Fig. 1: Pipeline stages are chained together into a 'production' handled by the DIRAC Transformation System.

Transformation System for CTA Monte Carlo simulations campaigns

- Several MC campaigns in 2013-2014
- Stable regimes of 4000-5000 jobs
- 6.7 M jobs executed
- 15 M replicas registered in the Catalog
- Transformation System in use since 2014

→ Efficient management of the productions

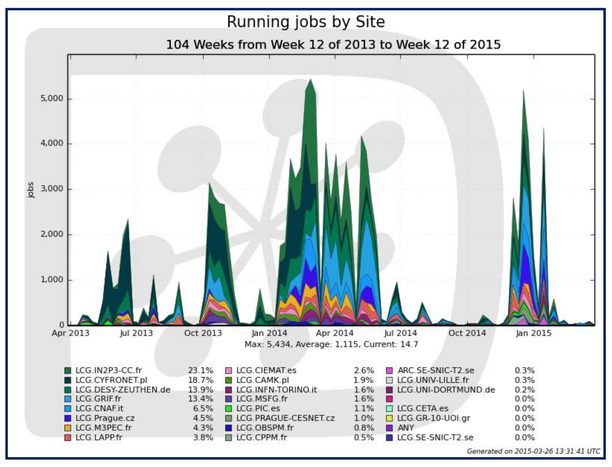


Fig. 2: Simulation productions during last 2 years.

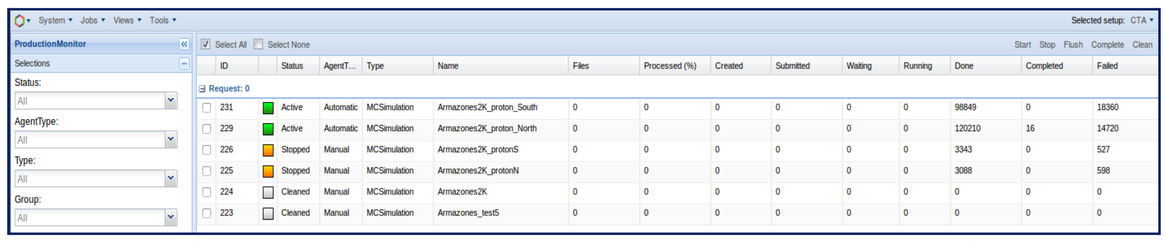


Fig. 3: Transformation System monitoring.

Towards a data-driven Production System

Current architecture:

- The Production Manager defines the transformations with meta-data conditions and 'plugins'
- The InputData Agent queries the Archive to obtain files to be 'transformed'
- Plugins group files into tasks according to desired criteria
- Tasks are created and submitted to the Workload Management System

Motivations for improvements:

- Querying the Archive may become a bottleneck
- Agents work in 'polling' mode
- Need to support chained transformations

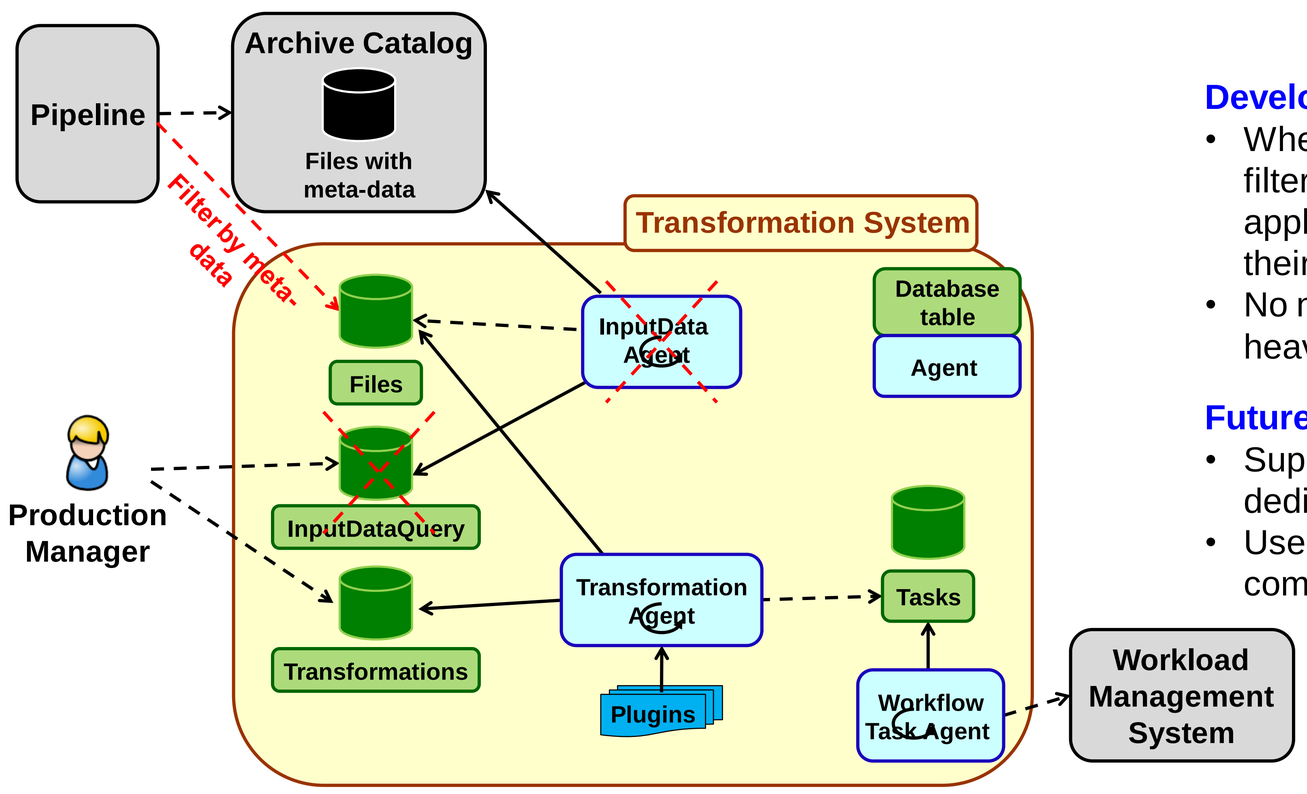


Fig. 4 : Simplified view of the DIRAC Transformation System architecture and of the proposed evolution in red.

Developments in progress:

- When new files are registered, a filter based on meta data is applied to send them on the fly to their matching transformation
- No need anymore to perform heavy queries of the Archive

Future improvements:

- Support chained transformations via dedicated meta-data
- Use Message Queue for the Agents complementary to polling

References

[1] M. Actis et al. (CTA Consortium), 2011, Experimental Astronomy, 32, 193

Acknowledgements: We gratefully acknowledge support from the agencies and organisations listed in this page: http://www.cta-observatory.org/?q=node/22

http://www.cta-observatory.org