LST0SA: Onsite processing pipeline for the CTA Large Size Telescope prototype

D. Morcuende[1], L. Saha[1], A. Baquero[1], J.L. Contreras[1], I. Aguado[1] and J.E. Ruiz[2] for the CTA LST project

[1]IPARCOS and EMFTEL Department, Universidad Complutense de Madrid, [2] Instituto de Astrofísica de Andalucía - CSIC

Introduction

The prototype of the Large Size Telescope (LST-1) [1] of the Cherenkov Telescope Array (CTA) [2], located at the Observatorio del Roque de Los Muchachos (ORM) on the Canary Island of La Palma, is presently going through its commissioning phase. A total of four LSTs, among other different-size telescopes, will operate together at ORM as part of the CTA North site.

Due to the large amount of daily recorded data, transferring the raw data through the network connection from La Palma to continental Europe in due time will be an issue for the LST. Therefore an LST On-Site Analysis (LST0SA) chain is being developed to perform the reduction of the raw data at the LST site.

Goals

- Process the data taken with the LST prototype.
- Provide low and intermediate level analysis products to the LST Collaboration.
- Data quality checks and help debugging potential problems.
- Track the provenance of the analysis products to ensure reproducibility.
- Fast offline high-level analysis on the site for targets of opportunity observations.

Computing infrastructure onsite

Acquisition, processing and analysis of the data.

- 55 computing nodes, each one with 32 cores.
- Several petabytes disk space.
- Batch queueing system, **SLURM** [4].
- Data acquisition rate ∼ 3 TB/hour of observation.

LST0SA

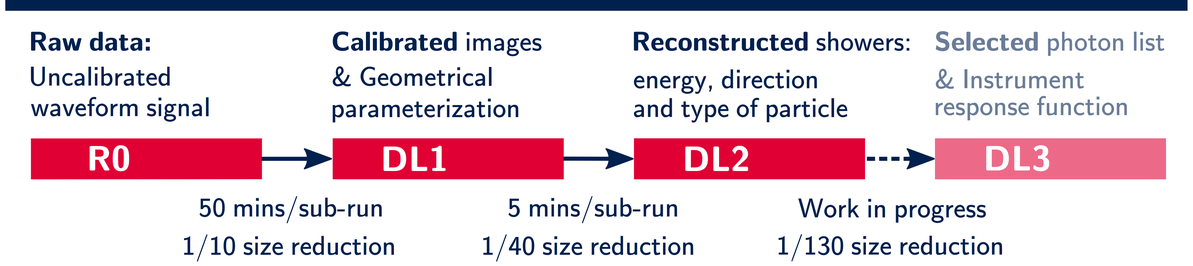
A set of python scripts connecting the different steps of the data reduction pipeline developed for the LST, **`lstchain`** [3]. It processes the data produced by the LST-1 in a semiautomatic way, producing high-level data and quality plots including detailed provenance logs which make the process reproducible. LST0SA splits the CPU-intensive first and second level data reduction steps of each observation run, in many sub-run jobs executed in parallel. Each observation run is normally composed of ∼10^2 sub-runs, having less than 10 seconds of data taking each. LST0SA produces calibrated data and image parameters **Data Level 1** and reconstructed showers **Data Level 2**.

The future CTA-North Array: CTA-N

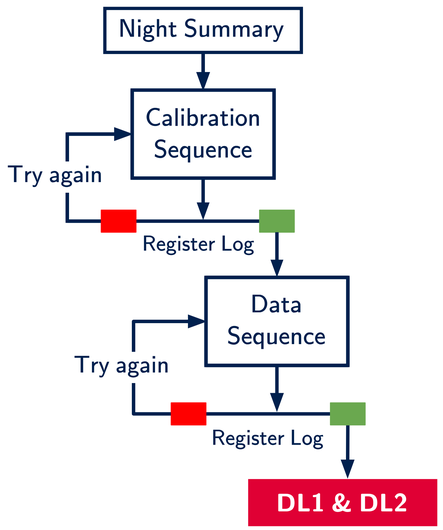
Artistic rendering of the future layout of the CTA-N array in the ORM observatory at La Palma. The two existing MAGIC telescopes point towards the sea, the LST prototype stands next to them.

The goal of LST0SA is to provide a solution for the onsite analysis of the whole CTA-North Array.

Data reduction steps



Work flow



Provenance

- **IVOA** provenance model with **W3C** syntax.
- Run-wise provenance products for each data level (`.json` serialization / `.pdf` graphs).
- Configuration files and input parameters kept run-wise to ensure **reproducibility**.
- **Next step:** use a **database** and develop a provenance **inspection query tool**.

Conclusion

A first version of the LST0SA pipeline is currently working at the Onsite processing system for the LST prototype with satisfactory results. It supports the commissioning of the telescope and the development of the analysis software on real data. It makes full use of the computing capacities of the Onsite computing cluster with a **high degree of automation**. Ongoing developments will add a higher automation degree and more flexibility.

References

[1] `http://lst1.iac.es/`

[2] `https://www.cta-observatory.org/`

[3] `https://github.com/cta-observatory/cta-lstchain`

[4] `https://slurm.schedmd.com/`

Acknowledgements

D. Morcuende acknowledges a predoctoral grant UCM-Harvard University (CT17/17–CT18/17) from Universidad Complutense de Madrid.

Contact Information

- Web: https://contrera.gitlab.io/lstosa/
- Email: dmorcuen@ucm.es