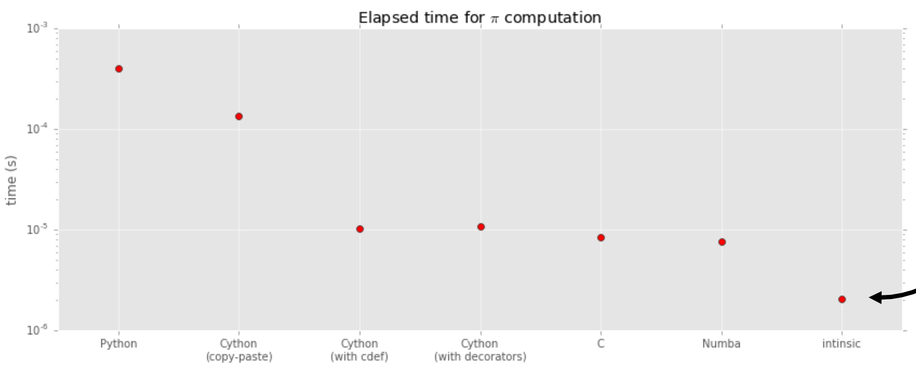
Application of HPC and vectorization solutions to Hillas-method reconstruction for Imaging Atmospheric Telescopes

Flash Me

The Cherenkov Telescope Array (CTA) represents the **new generation of Imaging Cherenkov Telescopes**. Its construction has already started and it is planned to reach its production phase in 2019. Its high sensitivity will generate about **4PB/yr of RAW data** which will require very efficient analysis software. CTA is thus – as other ESFRI experiments – definitely a BIG DATA project. The **ASTERICS-H2020** project aims at bringing together the ESFRI experiments to develop common solutions to the upcoming data management issues. The algorithms presented here have been developed under this project and are therefore available as open-source packages for the astrophysics community to use. Find it at https://gitlab.in2p3.fr/CTA-LAPP/CTA_Analysis (flash QRcode above).

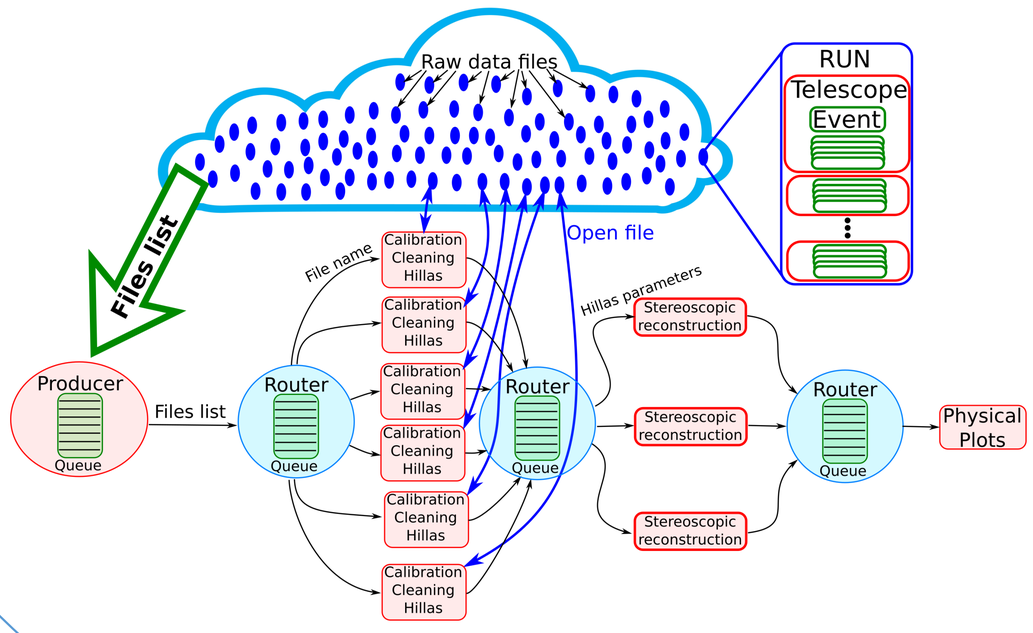
High Performance Computing Library

- Generated data format (in C++ and Python) optimized for vectorization
- Vectorized reduction and computation of first and second momenta (Hillas parameters)
- Picture matrix transformation to allow contiguous neighbours research and thus vectorize image cleaning
- Vectorized Python library (PLIB8 in CTA_Analysis) – comparison with other vectorized librairies

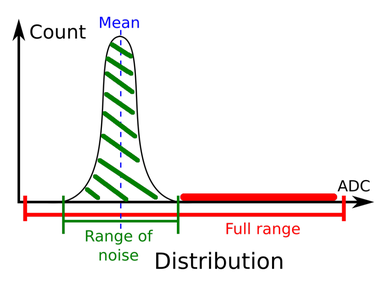


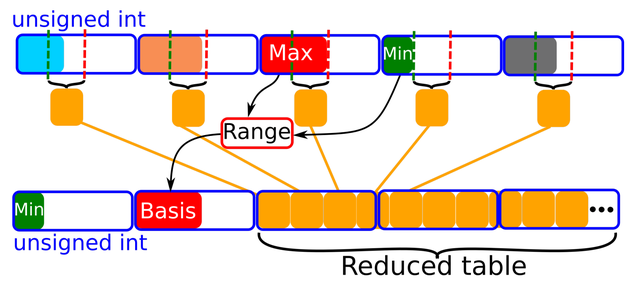
Workload management pipeline

ctapipe.flow is a Python implementation of the flow-based programming paradigm where applications are defined as networks of black-box components that exchange data across predefined connections.
ctapipe-flow executes processing modules in a sequential or multiprocessus environment with **automatic load**



Lossless Compression





Physics digitalized signals consist of integers. But the signal range (basis) is usually much smaller than the range of the values storable in an unsigned int. Knowing the min value and the basis of the signal, it can be stored as a single chunk of binary data (reduced table).

	Compression ratio	Time
LZMA (7z)	4.84	7 min 49s
Polynomial compression (ours)	3.74	3.7s
Polynomial compression + LZMA	**4.84**	**24.7s**

20 Monte-Carlo Simtel files (23Go of RAW data) can be*:

- compressed to 6.2Go in 4s

- analyzed in 9s

*(config: 2.27GHz, 2x12 GB RAM, 16 cores, SSE4 , cache L3 8 MB, L2 256 KB, L1 32 KB)

Authors: Vuillaume T*, Aubert P. Jacquemier J., Maurin G., Lamanna G. -- LAPP, CNRS, France
* thomas.vuillaume@lapp.in2p3.fr :: https://gitlab.in2p3.fr/CTA-LAPP/CTA_Analysis

Astronomy ESFRI & Research Infrastructure Cluster
ASTERICS - 653477